

# **BAB I**

## **PENDAHULUAN**

### **A. Latar Belakang**

Dalam era digitalisasi, perusahaan harus dapat mengambil keputusan yang tepat dan efektif untuk mencapai tujuan bisnis yang sesuai. Salah satu cara untuk membuat keputusan yang tepat adalah dengan melakukan analisis data. Data telah menjadi tulang punggung masyarakat modern, dalam setiap kegiatan yang berbasis digital.

Pengambilan keputusan efektif memerlukan informasi dari banyak sumber berbeda, seperti data historis, informasi eksternal, dan pendukung dari tim dan perusahaan lain. Pengambilan keputusan yang tepat dan efektif sangat penting bagi perusahaan, karena akan mempengaruhi masa depan dan keberhasilan perusahaan. Oleh karena itu, perusahaan harus memperhatikan tahapan pengambilan keputusan, seperti mengidentifikasi keputusan yang akan diambil, mengumpulkan informasi dan data pendukung, membuat alternatif pilihan, dan menganalisis data. Dengan melakukan analisis data, perusahaan dapat memperoleh kesimpulan yang lebih akurat dan andal, dan membuat keputusan yang tepat dan efektif.

Perkembangan era digital tidak terbatas pada satu rumpun ilmu saja, melainkan mencakup seluruh spektrum kegiatan manusia. Kemampuan untuk menganalisis data guna mendapatkan wawasan yang mendalam dan mengambil keputusan berbasis data menjadi salah satu pilar utama dalam mencapai kesuksesan dalam lingkungan yang semakin kompleks.

Dalam hal ini, data analytics muncul sebagai katalisator kunci yang memungkinkan perusahaan untuk menggali potensi dan peluang baru. Setiap aspek kehidupan bisnis, dari pemasaran hingga operasional, kini dapat ditingkatkan melalui pemahaman mendalam terhadap data yang dihasilkan.

Data analytics bukan hanya menjadi alat untuk meramalkan masa depan dan mengidentifikasi peluang inovatif, tetapi juga merupakan pondasi untuk evaluasi dan perbaikan dalam menghadapi tantangan saat ini.

Organisasi bisnis tidak hanya perlu mengumpulkan data, tetapi juga harus mampu memastikan keakuratannya. Oleh karena itu, kesinambungan pengumpulan, analisis, dan interpretasi data menjadi kritis dalam memastikan bahwa setiap langkah yang diambil oleh sebuah organisasi didasarkan pada informasi yang kuat dan relevan.

Peran data analis terdapat dalam menghadapi ketidakpastian dan kompleksitas, organisasi bisnis yang mampu mengoptimalkan potensi data dengan bijak akan memimpin jalan menuju kesuksesan jangka panjang dan relevansi di era digital ini.

Program Studi Independen Bersertifikat Data Analyst ini dihadirkan sebagai solusi konkrit untuk mengatasi permasalahan tersebut. Dengan menyelenggarakan program ini melalui platform Karier.mu, Sekolah.mu memberikan akses kepada para profesional dan pelajar untuk mengembangkan keterampilan analisis data yang sangat dicari oleh industri saat ini. Sertifikasi yang diberikan pada akhir program tidak hanya mencerminkan pencapaian akademis, tetapi juga memberikan pengakuan resmi terhadap keterampilan yang dimiliki peserta.

Melalui upaya ini, Sekolah.mu dan Karier.mu tidak hanya memberikan solusi untuk kebutuhan keterampilan di bidang analisis data, tetapi juga berkontribusi pada pengembangan SDM yang dapat mendukung pertumbuhan ekonomi Indonesia dalam era digital ini. Dengan menyelenggarakan program Studi Independen Bersertifikat Data Analyst, keduanya berperan sebagai agen perubahan dalam membangun generasi unggul yang siap menghadapi tantangan dan peluang di dunia kerja yang terus berkembang.

## B. Identitas MSIB

Karier.mu adalah platform peningkatan dan pengembangan karier dari Sekolah.mu yang berbasis teknologi dan berkolaborasi untuk menyediakan program pelatihan bagi semua orang. Tujuan utama Karier.mu yaitu menghubungkan semua orang kepada

mentor terbaik, untuk membantu mereka mewujudkan mimpi dan karier impian di masa depan. Karier.mu menyediakan berbagai program dengan mutu pelayanan yang berpihak kepada semua orang secara merata, serta menyediakan akses pelatihan terintegrasi digital dengan materi yang dirancang oleh ahli dari berbagai bidang.

MSIB yang diikuti dapat dikategorikan sebagai MSIB di bidang Data Analytics, yang diselenggarakan melalui platform Karier.mu. Kategori MSIB ini secara spesifik dirancang untuk membekali peserta dengan pengetahuan dan keterampilan terkini dalam analisis data, mencakup aspek-aspek kritis seperti mulai dari pengolahan data hingga interpretasi hasil yang dapat membantu perusahaan dalam pengambilan keputusan. Shortcourse ini memiliki tingkat level nasional, dengan fokus pada konteks dan kebutuhan pasar tenaga kerja di Indonesia. Meskipun berskala nasional, program ini tetap memenuhi standar internasional dalam penyampaian materi dan kurikulum, mengingat pentingnya mengikuti tren global dalam analisis data.

Dengan menghadirkan instruktur yang berpengalaman dalam industri Data Analytics, MSIB ini memberikan wawasan mendalam yang relevan dengan kondisi pasar tenaga kerja di Indonesia. Dengan demikian, program ini tidak hanya meningkatkan pemahaman peserta terhadap aspek teoritis analisis data, tetapi juga memberikan perspektif praktis yang dapat diaplikasikan dalam lingkungan kerja sehari-hari di tingkat nasional. Oleh karena itu, MSIB di bidang Data Analytics melalui Karier.mu ini dapat dianggap sebagai langkah yang relevan dan sesuai dengan kebutuhan dan perkembangan profesi di Indonesia.

a. Kategori

Program Studi Independen ini dilakukan selama 5 bulan dari tanggal 14 Agustus 2023 hingga 31 Desember 2023.

b. Level

Level program yang diselenggarakan Karier.mu ini termasuk dalam level nasional.

## **BAB II**

### **GAMBARAN MSIB**

#### **A. Penyelenggaraan**

PT. Semesta Integrasi Digital (PT.SID) adalah perusahaan yang bergerak dibidang pendidikan dan teknologi. Didirikan pada 2019 oleh Najeela Shihab dan dikembangkan oleh para profesional di bidang pendidikan dan teknologi. PT. SID bertujuan untuk mewujudkan akses pendidikan dan pengembangan talenta yang berkualitas untuk semua individu. Salah satu program yang dimiliki PT. Semesta Integrasi Digital (PT.SID) yaitu sekolah.mu. Sebuah lembaga pendukung pendidikan terkurasi yang menyediakan berbagai program pembelajaran terakreditasi mulai dari anak usia dini hingga sekolah menengah, guru, dan pengembangan profesional lainnya. Berkolaborasi dengan sekolah-sekolah mapan, lembaga kejuruan, dan pelaku industri, seluruh program di Sekolah.mu merupakan perpaduan pengalaman online dan offline yang dirancang untuk meningkatkan pembelajaran berkualitas bagi setiap peserta didik. Dalam hal ini, unit program tersebut menjadi penyelenggara shortcourse pada bidang data analyst.

#### **B. Lokasi**

PT Semesta Integrasi Digital berlokasi di Jl. Jeruk Purut Dalam No.33, RT.6/RW.3, Cilandak Tim., Kec. Ps. Minggu, Kota Jakarta Selatan, Daerah Khusus Ibukota Jakarta 12560. Namun, dalam pelaksanaan shortcourse data analyst yang disalurkan melalui program Magang dan Studi Independen Bersertifikat (MSIB) dilakukan dengan metode online pada platform Karier.mu yang dapat diakses kapanpun dan dimanapun.

#### **C. Waktu Pelaksanaan**

Program studi independent ini berjalan kurang lebih 5 bulan, sejak 14 Agustus 2023 hingga 31 Des 2023, dengan output proyek akhir. Setelah peserta menjalani rangkaian

acara hingga pada presentasi show case kemudian peserta melaksanakan graduation pada tanggal 22 Desember 2023.

Adapun rincian kegiatan pembelajaran selama 5 bulan tersebut yaitu:

Tabel 2. 1 Rincian Kegiatan Pembelajaran

Modul	Mulai	Selesai
Modul 1: Data Analytics Fundamental & Business Questions Formulation	16 Agustus 2023	03 September 2023
Modul 2: Prepare & Analyze Data with Google Spreadsheets	21 Agustus 2023	17 September 2023
Modul 3: Extracting Data with SQLite	11 September 2023	08 Oktober 2023
Modul 4: Programming with Python	02 Oktober 2023	29 Oktober 2023
Modul 5: Using Visualization to find Insights in Data	30 Oktober 2023	19 November 2023
Project Akhir	06 November 2023	10 Desember 2023
Showcase	18 Desember 2023	21 Desember 2023
Graduation	22 Desember 2023	22 Desember 2023

#### D. Ruang Lingkup MSIB

Ruang lingkup MSIB tersebut yaitu

1. *Google Colaboratory*, yang digunakan untuk menulis dan mengeksekusi kode program
2. Bahasa Pemrograman *Python*

3. *Tableau*, yang digunakan untuk melakukan visualisasi data
4. *DB Browser for SQLite*, yang digunakan untuk mengedit dan menjalankan *query* terhadap basis data *sqlite*.

### BAB III TAHAPAN PERSIAPAN MSIB

#### A. Rancangan Jadwal Persiapan MSIB

1. Pendaftaran : 18 April – 30 Juni 2023.
2. Pembukaan : 14 Agustus 2023.
3. Kegiatan Pembelajaran : 14 Agustus 2023 – 31 Desember 2023.
4. Penutupan : 31 Desember 2023

#### B. Jadwal Kegiatan MSIB

Jadwal kegiatan pembelajaran MSIB sebagai berikut :

1. 14 Agustus 2023 – 8 September 2023.

Tabel 3. 1 Jadwal Kegiatan 14 Agustus - 8 September

Pekan	Jumlah Hari	Durasi	Kegiatan pembelajaran
Minggu 1	3	6 jam	<ul style="list-style-type: none"><li>• Pembukaan kegiatan MBKM</li><li>• Pengenalan ruang lingkup MSIB</li></ul>
Minggu 2	3	6 jam	<ul style="list-style-type: none"><li>• <i>Introduction Data</i></li><li>• Proses analisis data</li><li>• <i>Data Analyst VS Data Scient</i></li></ul>
Minggu 3	3	6 jam	<ul style="list-style-type: none"><li>• <i>Data dalam dunis bisnis</i></li><li>• <i>Missing value, Duplicate, Typo</i></li><li>• <i>Outlier</i></li></ul>
Minggu 4	3	6 jam	<ul style="list-style-type: none"><li>• <i>Data Cleaning</i></li><li>• <i>Dasar Statistika</i></li><li>• <i>Testing</i></li></ul>

2. 9 September 2023 – 6 Oktober 2023.

Tabel 3. 2 Jadwal Kegiatan 9 September - 6 Oktober

Pekan	Jumlah Hari	Durasi	Kegiatan pembelajaran
Minggu 5	3	6 jam	<ul style="list-style-type: none"> <li>• <i>Correlation</i></li> <li>• <i>Regression</i></li> <li>• Introduction SQL</li> </ul>
Minggu 6	3	6 jam	<ul style="list-style-type: none"> <li>• <i>DDL dan DML</i></li> <li>• <i>Tipe data SQL</i></li> <li>• <i>Basiq SQL Syntak</i></li> </ul>
Minggu 7	3	6 jam	<ul style="list-style-type: none"> <li>• <i>Agregation in SQL</i></li> <li>• <i>Windows Function</i></li> <li>• <i>Case When</i></li> </ul>
Minggu 8	3	6 jam	<ul style="list-style-type: none"> <li>• <i>Penggabungan Tabel</i></li> <li>• <i>CTE</i></li> <li>• <i>Subquery</i></li> </ul>

3. 9 Oktober 2023 – 3 November 2023

Tabel 3. 3 Jadwal Kegiatan 9 Oktober - 3 November

Pekan	Jumlah Hari	Durasi	Kegiatan pembelajaran
Minggu 9	3	6 jam	<ul style="list-style-type: none"> <li>• <i>Introduction with Python</i></li> <li>• <i>Funtion, Conditional, Looping</i></li> <li>• <i>Preprocessing &amp; Cleaning data</i></li> </ul>
Minggu 10	3	6 jam	<ul style="list-style-type: none"> <li>• <i>EDA &amp; A/B Testing</i></li> <li>• <i>Cluster Analyst</i></li> <li>• <i>Correlation &amp; Regression analyst</i></li> </ul>
Minggu 11	3	6 jam	<ul style="list-style-type: none"> <li>• <i>Forecasting analyst</i></li> <li>• <i>Funnel analysis</i></li> <li>• <i>Insight Generation</i></li> </ul>
Minggu 12	3	6 jam	<ul style="list-style-type: none"> <li>• <i>Introduction tableau</i></li> <li>• <i>Chart type</i></li> <li>• <i>Aggregate function</i></li> </ul>



4. 6 November 2023 – 8 Desember 2023

Tabel 3. 4 Jadwal Kegiatan 6 November - 8 Desember.

Pekan	Jumlah Hari	Durasi	Kegiatan pembelajaran
Minggu 13	3	6 jam	<ul style="list-style-type: none"> <li>• Table calculations</li> <li>• <i>Parameter</i></li> <li>• <i>Dashboard dengan tebleau</i></li> </ul>
Minggu 14	3	6 jam	<ul style="list-style-type: none"> <li>• <i>Pentingnya visualisasi data</i></li> <li>• <i>Deck hasil analisis data</i></li> <li>• <i>Cara mempresentasikan data</i></li> </ul>
Minggu 15	3	6 jam	<i>Pengerjaan proyek akhir</i>
Minggu 16	3	6 jam	<i>Pengerjaan proyek akhir</i>

5. 11 Desember 2023 – 31 Desember 2023

Tabel 3. 5 Jadwal Kegiatan 11 Desember - 31 Desember

Pekan	Jumlah Hari	Durasi	Kegiatan pembelajaran
Minggu 17	3	6 jam	<i>Pengerjaan proyek akhir</i>
Minggu 18	3	6 jam	<i>Pengerjaan proyek akhir</i>
Minggu 19	3	6 jam	<i>Showcase</i>
Minggu 20	3	6 jam	<i>Graduation</i>

## **BAB IV**

### **HASIL PELAKSANAAN PRAKTIK MAGANG**

#### A. Hasil MSIB

##### 1. Proses MSIB

Pada program MBKM terdapat banyak cabang program yang disediakan. Salah satunya yaitu Magang dan Sudi Independen Bersertifikat (MSIB). Program ini hanya diperbolehkan memilih salah satu. Program yang diambil yaitu program studi independen. Program studi independent yaitu bagian dari program Kampus Merdeka yang bertujuan untuk memberikan kesempatan kepada mahasiswa untuk belajar dan mengembangkan diri melalui aktivitas di luar kelas perkuliahan, namun tetap diakui sebagai bagian dari perkuliahan. Program ini diperuntukan bagi mahasiswa yang ingin memperlengkapi dirinya dengan menguasai kompetensi spesifik dan praktis yang juga dicari oleh dunia usaha dunia industri. Studi independen yang diikuti memiliki jumlah konversi 20 SKS.

Program studi independen ini, setiap mahasiswa akan dikelompokkan kedalam suatu kelompok dan mendapatkan pembimbing pada setiap kelompok. Metode pembelajaran dilakukan dengan asynchronous dan synchronous. Metode asynchronous dilakukan dengan menggunakan app.karier.mu. Dalam LMS tersebut terdapat pretest, materi untuk setiap modulnya, terdapat quiz, terdapat ringkasan materi, praktik mandiri dan posttest. Sedangkan, untuk metode synchronous dilakukan dengan live session melakukan zoom bersama dengan expert. Pada live session setiap kelompok memiliki jadwal yang berbeda namun dengan frekuensi yang sama. Live session dilakukan sebanyak 2-3 kali pertemuan dalam seminggu namun tergantung modul yang sedang dipelajari. Waktu untuk live session diadakan pagi pukul 09.00-10.00, sore pukul 15.00-17.00, malam pukul 19.15-21.15. Adapun kompetensi yang dikembangkan:

##### 1. Programming with Python.

2. Prepare & Analyze Data with Google Spreadsheet.
3. Extracting Data with Big Query.
4. Data Analytics Essentials & Business Questions Formulation.
5. Using Visualization to find Insights in Data

Setiap modul memiliki quiz dan praktik mandiri yang wajib dikerjakan untuk setiap minggu. Selanjutnya, setelah peserta menyelesaikan setiap modulnya dari modul 1 hingga modul 5 beserta dengan tugas-tugas, kemudian peserta akan mendapatkan proyek akhir. Proyek akhir ini diawali dengan setiap peserta dalam satu kelompok dibagi kembali menjadi beberapa kelompok. Masing-masing kelompok terdapat 4-5 anggota kelompok. Kemudian diberikan sebuah isu nyata oleh tim karier.mu. Selama proses pengerjaan, setiap satu minggu sekali terdapat mentoring proyek yang dibimbing langsung oleh pembimbing kelompok. Kemudian, setelah proses pengerjaan selesai setiap kelompok akan show case yang mana dalam hal ini, berarti setiap kelompok menampilkan hasil pengerjaannya dihadapan tim kariermu dan juga juri.

Program studi independent ini berjalan kurang lebih 5 bulan, sejak 14 Agustus 2023 hingga 31 Des 2023, dengan output proyek akhir. Setelah peserta menjalani rangkaian acara hingga pada presentasi show case kemudian peserta melaksanakan graduation pada tanggal 22 Desember 2023.

## 2. Ringkasan Materi MSIB

### 1) Introduction Data

Data adalah sekumpulan fakta, statistik, atau informasi, yang diperoleh dari hasil pengukuran atau pengamatan suatu objek. Contoh data: catatan belanja bulanan, foto, video, atau lagu yang ada di smartphone. Big data adalah data yang ukurannya sangat besar, sifatnya kompleks, dan jumlahnya terus bertambah. setiap waktu dengan sangat cepat, sehingga sudah tidak mungkin untuk disimpan, dan diproses pakai metode atau alat komputasi tradisional.

Contoh: data-data browsing produk, akses halaman produk, add to cart, detail pembayaran dalam suatu situs e-commerce. Karakteristik Big Data (5V):

1. Volume atau ukuran data.

Contoh: Facebook punya sebanyak 2,8 miliar pengguna aktif bulanan.

2. Velocity atau kecepatan produksi data.

Contoh: Pada tahun 2018, setiap hari, seluruh orang di dunia, menghasilkan data sebesar 33 Zetabytes atau sama dengan 33 triliun Gigabytes.

3. Variety atau keragaman jenis data.

Menurut formatnya, keragaman data dapat dikategorikan sebagai berikut:

- Structured data: Structured Data (data terstruktur) adalah data yang siap digunakan atau dianalisa, formatnya tetap, dan dapat langsung disimpan dalam bentuk digital. Contoh: data pembelian customer, yang disimpan dalam format XLSX.
- Semi-structured data: Semi-Structured Data (data semi terstruktur) adalah data yang sudah diketahui informasi di dalamnya, namun format datanya campuran antara data terstruktur dan tidak terstruktur. Contoh: data hasil crawling Twitter dalam format JSON).
- Unstructured data: Unstructured Data (data tidak terstruktur) adalah data yang formatnya masih beragam sehingga perlu diolah dulu supaya bisa digunakan dan diketahui value-nya. Contoh: teks pada postingan blog, foto diambil dengan kamera smartphone, video YouTube, dan musik Spotify

4. Veracity alias tingkat kepastian atau keakuratan data. Kalau data yang kita pakai ternyata bias, sudah tidak valid, sumbernya tidak jelas, atau kondisi datanya masih kotor saat dianalisa, hasil atau insight yang diperoleh jelas dipertanyakan, sehingga keputusan yang diambil berdasarkan insights tersebut kemungkinan besar pasti salah

5. Nilai dari data itu sendiri. Data yang baik adalah data yang pada akhirnya memiliki nilai guna atau value.

## 2) Proses analisis data

Berikut merupakan langkah proses analisis data:

### 1. Definisikan tujuan dari proses analisis data

Tujuan analisis menjadi dasar penentuan konteks atau lingkup dari data yang akan digunakan. Tujuan analisis jadi tolak ukur apakah insights yang diperoleh menjawab permasalahan yang ingin dipecahkan.

### 2. Mengumpulkan data. Sumbernya bisa dari data internal atau eksternal dari perusahaan.

### 3. Menyiapkan data

Contoh yang harus diatasi saat pembersihan data:

- Nilai yang tidak valid
- Nilai yang ambigu Nilai yang kosong
- Data duplikat
- Nilai ekstrem

### 4. Analisis data

Menurut jenis datanya, teknik analisis data terbagi menjadi dua, yaitu teknik analisis data kualitatif, dan teknik analisis data kuantitatif.

#### A. KUANTITATIF Menganalisis data berupa angka-angka atau statistik.

Contoh data: Demografi penduduk, Skor Kredit, Harga property

#### B. KUALITATIF Menganalisis data non-numerik

Contoh data: Review produk ,Sejarah, Rekaman suara

### 5. Interpretasi temuan dari hasil analisis dan rekomendasi berdasarkan temuan. Hasil dari interpretasi data harus sejalan dan bisa mencapai tujuan dari proses analisis data yang sudah ditentukan sebelumnya

### 3) Data Analyst VS Data Scientist



Data Analyst vs Data Science		
<b>Deskripsi Pekerjaan</b>	<b>DATA ANALYST:</b> <ul style="list-style-type: none"> <li>● Menganalisis data</li> <li>● Mencari insights yang berguna dari data</li> <li>● Menyampaikan hasil analisis</li> <li>● Memberikan rekomendasi strategi bisnis</li> <li>● Membuat reports atau dashboards</li> </ul>	<b>DATA SCIENTIST:</b> <ul style="list-style-type: none"> <li>● Proses analisis data adalah bagian dari rangkaian proses yang lebih besar (modelling)</li> <li>● Menggunakan berbagai teknik statistika yang lebih kompleks, machine learning, atau algoritma-algoritma lainnya.</li> </ul>
<b>Keterampilan</b>	<b>Keterampilan untuk jadi Data Analyst:</b> <ul style="list-style-type: none"> <li>● Berkomunikasi dengan baik</li> <li>● Memahami konteks dan masalah bisnis</li> <li>● Menguasai ilmu statistika dasar</li> </ul>	<b>Keterampilan untuk jadi Data Scientist:</b> <ul style="list-style-type: none"> <li>● Berkomunikasi dengan baik</li> <li>● Memahami konteks dan masalah bisnis</li> <li>● Menguasai ilmu statistika</li> <li>● Menguasai ilmu matematika</li> <li>● Menguasai bahasa pemrograman</li> </ul>
<b>Bahasa Pemrograman &amp; Tools</b>	<b>Keterampilan untuk jadi Data Analyst:</b> <ul style="list-style-type: none"> <li>● Menguasai SQL untuk ekstraksi data</li> <li>● Microsoft Excel untuk eksplorasi data hingga analisis statistik deskriptif pada umumnya sudah memadai</li> <li>● Tableau untuk visualisasi data, membuat reports, dashboard</li> <li>● R dan Python untuk eksplorasi, analisis data dengan statistika dasar</li> </ul>	<ul style="list-style-type: none"> <li>● Menguasai SQL untuk ekstraksi data</li> <li>● Menggunakan R dan Python untuk eksplorasi data, analisis statistik lanjutan, dan machine learning untuk model</li> </ul>

Gambar 4. 1 Data Analyst Vs Data Science

### 4) Data dalam dunia bisnis

Dengan adanya data dan kemampuan buat analisis data, perusahaan bisa menggunakan hasil analisis tersebut sebagai acuan dalam mengambil keputusan dan menyusun strategi, berdasarkan bukti yang kuat dan sesuai

dengan kaidah keilmuan, sehingga kemungkinan terjadinya pengambilan keputusan atau tindakan yang salah makin minim. Diperlukan pemahaman terkait bussines matriks:

- Sebuah ukuran/indikator bisnis terukur yang digunakan untuk melacak, monitor, dan mengevaluasi berbagai proses bisnis sehingga perusahaan dapat menilai apakah proses bisnis / produk yang mereka miliki berjalan sukses atau gagal.
- Tujuan menggunakan metrik adalah:  
Mengkomunikasikan suatu kemajuan dari organisasi/perusahaan dalam jangka pendek/Panjang serta membantu para manajer untuk membuat keputusan yang lebih baik dan mengidentifikasi dari segi operasional bisnis manakah yang lebih efektif.  
Tiap unit bisnis di berbagai perusahaan belum tentu menggunakan metrik yang sama dalam mengukur kinerja, karena hal ini tergantung dari masing-masing kebutuhan, tujuan dan industri bisnis itu sendiri.
- Istilah terkait → KPI (Key Performance Indicator), OKR (Objective & Key Results)

#### 5) Missing value, Duplicate, Typo

Missing value merujuk pada nilai yang absen atau tidak ada dalam dataset. Missing value dapat terjadi karena berbagai alasan, termasuk kesalahan manusia, kegagalan perangkat, atau sifat data yang tidak lengkap. Missing value dapat menyebabkan bias dalam analisis data dan mengurangi keakuratan model yang dibangun atas dasar data tersebut. Untuk menangani missing value dapat dilakukan dengan:

- Menghapus baris atau kolom dengan banyak missing value.
- Mengisi missing value dengan nilai rata-rata, median, atau modus dari kolom yang bersangkutan.

Duplicated (data duplikat) merupakan pada keberadaan data yang identik atau sangat mirip dalam dataset. Data duplikat bisa muncul karena kesalahan dalam pengumpulan data atau duplikasi yang sengaja dilakukan. Data duplikat dapat mengganggu analisis data, memperbesar sampel yang seharusnya lebih kecil, dan menyebabkan bias dalam estimasi statistik. Untuk menangani hal tersebut dapat dilakukan:

- Mengidentifikasi dan menghapus data duplikat dari dataset.
- Memeriksa keaslian data saat pengumpulan data baru.

*Typo* (kesalahan ketik) adalah kesalahan dalam penulisan atau input data, yang bisa berupa perbedaan huruf, angka, atau simbol dari nilai yang seharusnya dimasukkan. *Typo* umumnya terjadi karena kesalahan manusia saat menginput data. *Typo* dapat menyebabkan ketidakakuratan dalam analisis data dan menghasilkan kesimpulan yang salah. Untuk menangani hal tersebut dapat dilakukan dengan melakukan pemeriksaan dan pembersihan data secara rutin untuk mendeteksi kesalahan ketik.

#### 6) Outlier

Outlier adalah titik data yang secara signifikan berbeda dari sebagian besar data lain dalam dataset. Outlier bisa muncul karena kesalahan pengukuran, anomali alam, atau kejadian langka. Outlier dapat menyebabkan kesalahan dalam analisis statistik dan mengganggu kinerja model prediktif.

Untuk menangani outlier dapat menggunakan pemeriksaan apakah outlier merupakan data yang valid atau kesalahan. Dapat juga dilakukan dengan menghapus outlier jika dapat diidentifikasi sebagai kesalahan.

#### 7) Data Cleaning

Data cleaning adalah proses memastikan keakuratan, konsistensi, dan kegunaan data dalam kumpulan data. Prosesnya yang dilakukan adalah mendeteksi kesalahan data atau data yang corrupt dan memperbaiki atau menghapus data sesuai kebutuhan. Mengapa Data Cleaning Harus Dilakukan?



1. Menghilangkan kesalahan dan inkonsistensi yang muncul saat beberapa data sources dikumpulkan dalam satu dataset.
2. Meningkatkan efisiensi kerja karena proses ini akan memudahkan pengolahan data untuk menemukan apa yang dibutuhkan dari data.
3. Tingkat error yang lebih rendah juga akan mendatangkan kepuasan pelanggan dan mengurangi beban kerja tim.
4. Membantu memetakan beberapa fungsi data yang berbeda. Proses ini juga akan membuat untuk lebih mengenal kegunaan data dan mempelajari asalnya.

#### Tahapan Cleaning Data

1. Mendeteksi kesalahan data
2. Melakukan cek duplikasi data
3. Memperbaiki kesalahan struktur data
4. Filter outlier yang tidak diinginkan
5. Tangani data yang hilang

#### 8) Dasar Statistika

##### - Analisis Statistik Deskriptif

Analisis statistik deskriptif digunakan untuk mendeskripsikan kumpulan data. Analisis ini akan menggambarkan karakteristik dasar, informasi, dan ringkasan data dengan cara rasional. Metode ini mendeskripsikan data secara kuantitatif. Analisis statistik deskriptif melibatkan berbagai proses seperti tabulasi, mean, median, range, variasi, standar deviasi, hingga analisis deret waktu.

##### - Analisis statistika inferensial

Statistika inferensial adalah cabang dari statistika yang berkaitan dengan membuat kesimpulan atau inferensi tentang populasi berdasarkan sampel data yang diambil dari populasi tersebut. Tujuan utama dari

statistika inferensial adalah untuk membuat generalisasi, membuat estimasi, menguji hipotesis, dan melakukan prediksi tentang populasi berdasarkan informasi yang terdapat dalam sampel data.

## 9) Testing

➤ A/B Testing melibatkan pengambilan sampel secara acak dan membaginya menjadi 2 kelompok (control dan treatment).

a. Variabel control: tidak diberikan treatment apapun → tidak diberikan voucher.

b. Variabel treatment: diberikan treatment → diberikan voucher.

Manfaat A/B testing:

- Membantu menyelesaikan masalah korelasi dan bias seleksi. Korelasi (hubungan 2 arah).
- Memungkinkan kita untuk menilai hubungan sebab akibat antara variabel.
- Mudah diimplementasikan dalam aplikasi web/ platform digital.
- Menggunakan data daripada opini untuk melawan HIPPO (Highest Paid Person's Opinion).

Konsep Dasar A/B Testing

- Null Hypothesis: Tidak ada hubungan antara 2 kelompok.
- Signifikansi Statistik: Seberapa besar kepercayaan kita terhadap sebuah experiment (biasanya dalam bentuk alpha). Signifikansi Statistik digunakan untuk dibandingkan dengan p-value
- P-value: Probabilitas yang menyatakan bahwa kejadian yang kita amati tidak terjadi secara kebetulan.

➤ One sample t-test

Simple t-Test: Apakah ada perbedaan antara sebuah grup dengan populasi

→ Apakah meja yang diproduksi memiliki berat rata-rata 2 kg.

➤ Two Sample t-test

- a. t-Test untuk independent samples  
t-Test untuk independent samples: Apakah ada perbedaan antara 2 grup  
→ Apakah ada perbedaan antara gaji grup A dan grup B. Sampel yang digunakan tidak ada hubungan.
- b. t-Test untuk dependent samples  
t-Test untuk dependent samples: Pengukuran berulang dengan orang sama tetapi dengan waktu yang berbeda.

#### 10) Correlation

Metode statistik yang digunakan untuk mengukur hubungan antara dua atau lebih variabel. Korelasi mengukur sejauh mana dua variabel bergerak bersama-sama atau berlawanan. Korelasi dapat memberikan wawasan tentang sejauh mana perubahan dalam satu variabel berkaitan dengan perubahan dalam variabel lainnya. Jenis-jenis uji korelasi:

- Korelasi Pearson  
Jenis korelasi yang paling umum digunakan, dan mengukur hubungan linier antara dua variabel kontinu. Korelasi Pearson menghasilkan koefisien korelasi antara -1 dan 1. Nilai -1 menunjukkan hubungan negatif sempurna, nilai 1 menunjukkan hubungan positif sempurna, dan nilai 0 menunjukkan tidak ada hubungan linier antara dua variabel.
- Korelasi Spearman  
Jenis korelasi ini digunakan ketika data tidak terdistribusi normal atau ketika ada keraguan tentang keberlanjutan hubungan linier antara dua variabel. Korelasi Spearman menggunakan peringkat data yang diurutkan dan mengukur hubungan monotone antara variabel-variabel tersebut. Dalam analisis korelasi Spearman, langkah pertama adalah menghitung korelasi antara dua variabel yang memiliki data ordinal atau interval.
- - Korelasi Kendall

Seperti Korelasi Spearman, korelasi ini juga mengukur hubungan monotone antara variabel-variabel. Korelasi Kendall menggunakan peringkat data dan menghitung jumlah pasangan data yang sesuai (concordant) dan yang tidak sesuai (discordant) untuk mengukur korelasi. Korelasi Kendall fokus pada pengukuran sejauh mana dua variabel memiliki urutan yang serupa atau berbeda dalam data. Ini adalah alat yang berguna ketika data tidak memenuhi asumsi distribusi normal dan juga ketika data diukur dengan skala ordinal.

## 11) Regression

- Linear Regression

Suatu metode statistik yang digunakan untuk memodelkan hubungan linier antara dua atau lebih variabel. Linear regresi sering digunakan untuk memprediksi nilai dari satu variabel (variabel dependen) berdasarkan nilai dari satu atau lebih variabel lainnya (variabel independen). Tujuan utama dari linear regresi adalah menemukan garis atau permukaan yang paling baik menyesuaikan data, sehingga kita dapat menggunakannya untuk membuat prediksi atau analisis.

- Simple Linear Regression

Metode statistik yang digunakan untuk memodelkan hubungan linier antara dua variabel, yaitu satu variabel dependen (variabel yang ingin diprediksi) dan satu variabel independen (variabel prediktor). Dalam regresi linear sederhana, kita mencoba menemukan garis lurus terbaik yang menggambarkan hubungan antara kedua variabel tersebut.

- Multiple Regression Linear Berganda

Metode statistik yang digunakan untuk memodelkan hubungan antara satu variabel dependen (variabel yang ingin diprediksi) dan dua atau lebih variabel independen (variabel prediktor). Ini adalah pengembangan dari

regresi linear sederhana yang memungkinkan Anda untuk mempertimbangkan pengaruh beberapa variabel independen terhadap variabel dependen.

## 12) Introduction SQL

Menyimpan data di relational database mirip seperti menyimpan data di komputer lokal kita. Contoh penyimpanan data di Database Management System (DBMS) PostgreSQL. Untuk mengakses data dari database, dibutuhkan perintah yang dimengerti aplikasi mengelola database, yaitu Structured Query Language (SQL). Komponen dan terminologi RDBMS.

- Tabel adalah struktur yang berisi record dan field Tabel merepresentasikan suatu subjek yang spesifik (misal customer, student, album).
- Field atau atribut atau kolom merepresentasikan karakteristik suatu data (tinggi siswa, nama customer).
- Record adalah baris data di tabel, berisi nilai dari tiap field.
- Key adalah penanda atau identifier di dalam sebuah tabel.
  - Primary key : berfungsi untuk mengidentifikasi setiap record atau baris. Sifatnya unik dan berbeda antar baris.
  - Foreign key : menunjukkan relasi antara dua tabel. Nilai foreign key di suatu tabel harus sama dengan nilai primary key di tabel lain yang berhubungan
- Relationship: hubungan antara dua tabel (1-1, 1-many, many-many)
- Constraint : limitasi untuk nilai yang diperbolehkan diisi kesebuah kolom

## 13) DDL dan DML

Data Definition Language (DDL)

- Perintah-perintah yang digunakan untuk membuat, menghapus maupun memodifikasi struktur tabel di database kita.
- Perintah dasar DDL : CREATE,ALTER,DROP

Data Manipulation Language (DML)

- Perintah - perintah yang digunakan untuk memasukkan dan memodifikasi data di dalam database.
- Perintah dasar DDL : INSERT,UPDATE,DELETE

#### 14) Tipe data di SQL

Saat membuat suatu tabel, kita perlu menentukan tipe data untuk tiap kolom/field di tabel kita. Beberapa tipe data yang sering digunakan:

- String : tipe data yang dapat menyimpan kombinasi huruf, angka dan special character. Contoh tipe data string di database adalah VARCHAR , CHAR atau TEXT.
- Number : tipe data yang hanya menyimpan angka. Contohnya adalah INTEGER dan NUMERIC.
- Date and Time : tipe data yang menyimpan tanggal dan waktu. Contohnya DATE dan TIMESTAMP.
- Boolean : tipe data yang menyimpan nilai True/False atau 0/1.
- Composite : gabungan dari beberapa tipe data

#### 15) Basic SQL Syntax

##### ➤ SELECT & FROM Syntax

SELECT merupakan perintah menampilkan maupun mengambil sebuah data pada tabel. Sedangkan FROM merupakan syntax untuk menunjukkan table yang digunakan untuk mengekstraksi data. Contoh :

```
SELECT kolom1
, kolom2
, kolom3
FROM table as t1
```

##### ➤ LIMIT Syntax

LIMIT merupakan syntax yang digunakan untuk membatasi jumlah data yang diambil ketika menjalankan perintah. Contoh :

```
SELECT kolom1
```

```
, kolom2  
, kolom3  
FROM table as t1  
LIMIT 50
```

➤ ORDER BY syntax

ORDER BY merupakan syntax yang digunakan untuk mengurutkan tampilan dari hasil query/ekstraksi data dengan kolom tertentu. Contoh :

```
SELECT kolom1  
, kolom2  
, kolom3
```

➤ WHERE Syntax

WHERE merupakan syntax yang digunakan untuk melakukan filter kepada data berdasarkan kondisi tertentu. Contoh :

```
SELECT kolom1  
, kolom2  
, kolom3  
FROM table as t1  
WHERE kolom3 > 1  
ORDER BY kolom1  
LIMIT 50
```

## 16) Aggregation in SQL

Agregasi SQL adalah fungsi untuk mengkalkulasi sekumpulan nilai. Hal ini dilakukan dengan bantuan fungsi agregat, seperti SUM, COUNT, MIN, MAX, dan AVG.

### 1. Count Function

Fungsi count berguna untuk menghitung berapa banyak data yang ingin kita hitung.

### 2. Sum Function

Fungsi sum digunakan untuk menjumlahkan data-data numerik.

### 3. Avg Function

Fungsi avg digunakan untuk menghitung rata-rata angkanya.

### 4. Min Max Function

Fungsi MIN digunakan untuk menghitung atau menampilkan nilai terkecil, sedangkan MAX kebalikannya.

### 5. Group By Function

Group by digunakan untuk menghitung menghitung agregasi data dalam suatu kategori atau grup tertentu.

### 6. Having Function

Having mempunyai fungsi yang mirip dengan WHERE, namun having digunakan apabila kita ingin memfilter suatu hasil agregasi

## 17) Windows Function

Window Functions merupakan sebuah fungsi yang digunakan untuk melakukan komputasi kepada nilai-nilai yang ada dalam tabel tanpa harus mengelompokkan baris menjadi satu output, seperti halnya ketika melakukan komputasi pada umumnya. Terdapat beberapa parameter yang digunakan untuk melakukan kalkulasi seperti :

1. SUM, AVG, COUNT
2. ROW\_NUMBER, RANK, DENSE\_RANK

Berikut adalah contoh syntax window functions yang pada umumnya digunakan:

```
SELECT calculation (column1) OVER (PARTITION BY column 2 ORDER BY column3) FROM table 1
```

Keterangan:

- Calculation: Kalkulasi yang akan dilakukan, biasanya menggunakan parameter SUM, AVG,COUNT, ROW\_NUMBER, RANK, DENSE\_RANK dan sebagainya .



- Over Klausa OVER menentukan apakah hasil kalkulasi akan ditampilkan berdasarkan seluruh baris atau beberapa baris tertentu. Apabila klausa ini dikosongkan OVER(), maka kalkulasi akan dilakukan berdasarkan seluruh baris yang ada.
- Partition By: Syntax untuk menentukan apakah kalkulasi akan dibagi berdasarkan kolom tertentu atau tidak.
- Order By: Syntax untuk menentukan apakah hasil kalkulasi akan diurutkan berdasarkan kolom tertentu atau tidak

#### 18) Case When

CASE WHEN atau yang lebih sering disebut sebagai statement CASE merupakan perintah untuk mengembalikan suatu value tertentu ketika terdapat kondisi yang terpenuhi. Fungsi ini sangat mirip dengan fungsi IF-THEN-ELSE pada Microsoft Excel, yang mana akan mengembalikan suatu value apabila kondisi yang ditentukan pada klausa IF terpenuhi, dan mengembalikan value lain yang dituliskan pada klausa ELSE apabila kondisi tersebut tidak terpenuhi.

Berikut merupakan contoh syntax case statement:

```

CASE expr
WHEN expr_to_match THEN result
[ ... ]
De such a fer.mu
[ ELSE else_result ]
END

```

Keterangan:

- CASE: Statement untuk memulai syntax CASE.
- WHEN: Statement yang berfungsi seperti IF, untuk menandakan kondisi yang harus terpenuhi.
- THEN Statement yang berfungsi sebagai THEN, untuk menandakan value apa yang akan dikembalikan apabila kondisi pada WHEN terpenuhi.

- ELSE: Bersifat optional yang mana akan memberikan perintah untuk mengembalikan value apabila seluruh kondisi pada statement WHEN/THEN tidak terpenuhi.
- END: Statement untuk mengakhiri syntax CASE. Statement ini bersifat wajib untuk dituliskan.

## 19) Penggabungan tabel

Join merupakan sebuah perintah yang digunakan untuk menggabungkan baris dari 2 tabel atau lebih yang didasarkan kepada kolom yang memiliki nilai yang sama di antara tabel-tabel tersebut. Ada beberapa jenis join, diantaranya yaitu :

- Inner Join
- Left Join
- Right Join
- Full Outer Join



Gambar 4. 2 Jenis-jenis join

## 20) CTE

*Common Table Expressions* (CTE) adalah tabel sementara yang dapat di buat dalam pernyataan SQL. CTE sangat berguna ketika perlu melakukan operasi kompleks seperti penggabungan, agregasi, atau rekursi. Pernyataan CTE dimulai dengan klausa WITH, diikuti oleh nama CTE dan definisinya. Definisi CTE berisi pernyataan SQL yang menghasilkan satu set hasil yang akan digunakan dalam

pernyataan utama. CTE dapat digunakan dalam klausa SELECT, FROM, atau WHERE pernyataan utama.

```
WITH CTE_Nama AS (SELECT kolom1, kolom2 FROM nama_tabel1),
```

```
CTE_Nama2 AS (SELECT kolom1, kolom2 FROM nama_tabel2 SELECT * FROM  
CTE_Nama1
```

Dengan memahami penggunaan CTE, dapat membuat pernyataan SQL yang lebih bersih, lebih terstruktur, dan lebih mudah dibaca. CTE sangat berguna dalam pernyataan SQL yang kompleks atau yang melibatkan operasi yang perlu diulang.

## 21) Subquery

*Subquery*, juga dikenal sebagai subselect atau nested query, adalah pernyataan SQL yang tertanam di dalam pernyataan SQL lainnya. Subquery digunakan untuk menggali data lebih lanjut dan digunakan dalam klausa WHERE, FROM, atau SELECT.

- Subquery dalam Klausa WHERE

Subquery dalam klausa WHERE digunakan untuk memfilter hasil dari pernyataan utama berdasarkan kondisi yang berasal dari subquery.

- Subquery dalam Klausa FROM

Subquery dalam klausa FROM digunakan untuk membuat tabel sementara yang dapat digunakan dalam pemrosesan utama.

- Subquery dalam Klausa SELECT

Subquery dalam klausa SELECT digunakan untuk menghasilkan nilai yang akan dimasukkan ke dalam hasil akhir pernyataan SQL.

## 22) Introduction with Python

➤ Variabel dalam Python

Variabel dalam Python adalah seperti wadah atau tempat penyimpanan yang digunakan untuk menampung data atau nilai. Kamu dapat memberi

nama pada variabel sesuai kebutuhan mu dan kemudian menyimpan nilai di dalamnya.

➤ Contoh jenis-jenis Operator pada Python

Operator Aritmetika:

+ (Penambahan)

- (Pengurangan)

\* (Perkalian)

/ (Pembagian)

% (Modulus, sisa pembagian)

\*\* (Pangkat)

// (Pembagian bulat)

Operator Perbandingan:

== (Sama dengan)

!= (Tidak sama dengan)

< (Kurang dari)

> (Lebih dari)

<= (Kurang dari atau sama dengan)

>= (Lebih dari atau sama dengan)

➤ List

Dalam bahasa pemrograman Python, struktur data yang paling dasar adalah urutan atau lists. Setiap elemen-elemen berurutan akan diberi nomor posisi atau indeksinya. Indeks pertama dalam list adalah nol, indeks kedua adalah satu dan seterusnya.

➤ Method

Method adalah salah satu hal yang sangat penting dalam python. Fungsi yang dimiliki oleh suatu object dinamakan method. Contohnya kita membuat variabel bertipe string, bertipe list, bertipe numeric dan lain sebagainya, itu semua adalah object, dan setiap object memiliki fungsi

yang hanya bisa digunakan oleh object tersebut. Sebagai contoh, object string memiliki method uppercase, dimana method upper ini tidak bisa digunakan oleh object lain seperti list. Tapi list juga memiliki fungsi seperti index, yang dimana fungsi index ini tidak bisa digunakan oleh data dengan tipe objek lain seperti string contohnya.

### 23) Funtion, Conditional, Looping

#### ➤ Function

Function adalah cara menerapkan abstraksi dan dekomposisi di bahasa python. Cara kerjanya adalah function dapat menerima input dan menghasilkan output saat fungsi dipanggil, tanpa user yang menggunakan function tersebut perlu tahu secara detail bagaimana kode tersebut bekerja. Membuat Fungsi dalam Python suatu fungsi didefinisikan menggunakan kata kunci def

contoh

```
def my_function():  
    print("Hello from a function")
```

#### ➤ Conditional

Conditional statement merupakan sebuah handling pada bahasa pemrograman yang berfungsi untuk menyelesaikan logika bercabang ataupun pengambilan keputusan. Python pun memiliki beberapa bentuk conditional statement yang sering dijumpai, antara lain:

- if statement
- if-else statement
- if-elif-else statement

#### ➤ Looping

Looping adalah melakukan pengulangan / hal yang repetitif sampai suatu kondisi

terpenuhi atau tercapai.

Contoh :

- Menulis teks 'saya tidak akan mencontek lagi' 50 kali
- Mengirim email berdasarkan sebuah list alamat email di google sheet
- Memesan pizza secara online selama limit kartu kredit kita masih ada

Dalam python terdapat 2 jenis looping: while loop dan for loop

#### 24) Preprocessing & Cleaning data

##### ➤ Explorasi data awal

- `pandas.read_csv(filepath)` : kita dapat menggunakan fungsi ini untuk membuat suatu dataframe dari sebuah file csv, dengan memberikan input lokasi file kita berada
- `dataframe.head(jumlah_row)` : menampilkan N-baris teratas dari dataframe kita, tergantung dari input jumlah\_row
- `dataframe.info()` : menampilkan informasi untuk tiap kolom dari dataframe kita (tipe data, jumlah baris yang tidak null)
- `dataframe.shape` : menampilkan jumlah baris dan kolom di dataframe
- `series.value_counts()` : berguna untuk menampilkan nilai unik di kolom kategorikal dan jumlah tiap nilai unik tersebut

##### ➤ Mengatasi missing value dan duplikat

- Kita dapat menggunakan fungsi `isna()` dan `sum()` untuk melihat berapa jumlah missing value di tiap kolom di dataframe kita
- Kemudian kita bisa gunakan fungsi `dataframe/series.fillna(value)` untuk mengisi missing value di dataframe ataupun series/kolom tertentu. Fungsi ini memiliki argumen wajib value yaitu nilai untuk menggantikan missing value
- Fungsi `dataframe.duplicated()` ditambah dengan `sum()` akan membantu kita melihat apakah ada data duplikat di dataframe kita. Kita dapat menghapus semua duplikat dengan fungsi `drop_duplicates()`

##### ➤ Hapus (drop) serta penamaan ulang (rename) kolom dan baris

- Kita dapat gunakan `dataframe.drop()` untuk menghapus baris - baris data yang kita rasa tidak diperlukan
- Gunakan fungsi `dataframe.rename()` untuk memperbaiki nama kolom di dataframe kita
- Mendeteksi outlier dengan Interquartile Range (IQR)

## 25) EDA & A/B Testing

EDA merupakan suatu pendekatan dalam statistika dan analisis data yang digunakan untuk memahami dataset sebelum melakukan analisis statistik yang lebih mendalam atau pemodelan data. EDA bertujuan untuk mengidentifikasi pola, tren, anomali, dan karakteristik penting dalam data tanpa membuat asumsi sebelumnya tentang hasil yang diinginkan. Tujuan dari pengujian hipotesis adalah membuktikan suatu pernyataan umum dengan menggunakan data - data sampel yang ada. Prinsip yang digunakan adalah praduga tak bersalah (Presumption of Innocence)

Pengujian Hipotesis Dalam Statistika

Hipotesis 0 yang tadi kita sebutkan dalam statistika biasa disebut dengan Hipotesis Nol/ Null Hypothesis ( $H_0$ )

Null hypothesis ( $H_0$ )

Sebuah hipotesis yang berkontradiksi dari pernyataan yang akan diuji. Sederhananya hipotesis yang bersifat umum

Sedangkan Hipotesis 1 dalam statistika biasa disebut dengan Hipotesis Alternatif /Alternative Hypothesis ( $H_1/H_a$ )

Alternative Hypothesis ( $H_1$ )

Hipotesis yang ingin kita uji. Bersifat spesifik

Metrics untuk mengukur resiko kesalahan menolak  $H_0$  (error type 1) dinamakan P-Value (Probability Value).

P-Value ini berbentuk prosentase / peluang yang menggambarkan resiko ketika menolak  $H_0$  / menerima  $H_1$ .

Contoh :

Besarnya resiko kesalahan menyatakan bahwa Ali pencuri adalah 1%.

Cara Mengambil Keputusan

Kondisi pengambilan keputusan dibagi menjadi 2:

1. Jika P-value  $< \alpha$ , maka kita akan mengambil H1
2. Jika P-Value  $> \alpha$ , maka kita akan mengambil H0

Jenis-jenis pengujian statistika sangat banyak

1. T-Test
2. ANOVA
3. Pearson Correlation Test
4. Chi-Square
5. Wilcoxon Rank-Sum Test
6. Sign Test
7. e.t.c

## 26) Cluster Analyst

Clustering merupakan sebuah teknik yang digunakan untuk menemukan sub-kelompok atau kluster. Prinsip dari clustering adalah data-data yang termasuk dalam kluster yang sama “diharapkan” memiliki kesamaan satu dengan yang lainnya, sedangkan data-data yang ada dalam kluster yang berbeda cukup berbeda dari satu kluster dengan cluster yang lainnya. Meskipun tanpa label, clustering memanfaatkan “karakter/pola” pada data bahwa data yang “mirip” akan berdekatan

Proses standar dari sebuah algoritma clustering adalah sebagai berikut:

feature extraction and selection, ekstrak dan tentukan fitur yang paling representatif dari dataset clustering algorithm design, desain algoritma clustering sesuai dengan karakteristik dari masalah result evaluation, evaluasi hasil clustering result explanation, beri penjelasan praktis tentang hasil yang diperoleh dari clustering



Secara umum, terdapat 4 pendekatan clustering yang sering digunakan:

- centroid-based clustering

Centroid-based clustering mengelompokkan data berdasarkan sebuah centroid yang merupakan “pusat” dari sebuah cluster.

- density-based clustering

Density-based clustering akan membentuk cluster dari data-data yang memiliki kepadatan yang tinggi

- distribution-based clustering

Metode ini bekerja dengan cara mengasumsikan bahwa data tersusun atas beberapa distribusi, misalkan gaussian distribution

- hierarchical-clustering

Sesuai namanya, cluster yang dibentuk tersusun atas cluster kecil lain yang jika divisualisasikan akan berbentuk sebuah pohon hirarki

Metode ini sangat cocok digunakan pada data yang memiliki karakteristik hirarki, seperti taksonomi, user journey dan lainnya.

## 27) Correlation & Regression analyst

Regresi linier adalah suatu metode yang digunakan untuk menyatakan pola hubungan antara variabel respon dengan variabel prediktor.

Tujuannya adalah mengembangkan model statistik yang dapat memprediksi nilai variabel dependen (respon) berdasarkan nilai variabel independen (penjelas). Bila variabel predictor berjumlah lebih dari satu sehingga digunakan analisis regresi linier berganda.

Dalam korelasi, kedua variabel diperlakukan sama. Dalam regresi, satu variabel dianggap variabel independent X (prediktor) dan variabel dependen Y (respon atau hasil).

Ada beberapa manfaat menggunakan analisis regresi:

- Menunjukkan adanya hubungan yang signifikan antara variabel dependen dan variabel independen.

- Menunjukkan kekuatan pengaruh beberapa variabel independen terhadap variabel dependen.
- Analisis regresi juga untuk membandingkan pengaruh variabel yang diukur pada skala yang berbeda, seperti pengaruh perubahan harga dan jumlah kegiatan promosi.

Manfaat ini membantu peneliti pasar/analisis data/ilmuwan data untuk menghilangkan dan mengevaluasi kumpulan variabel terbaik yang akan digunakan untuk membangun model prediktif.

Simple linear regression diantaranya :

- regresi linier sederhana
- standart error regresi linier
- koefisien determinasi
- uji parsial regresi sederhana

Multiple linear regression diantaranya :

- regresi linier berganda
- modelling

Optimasi Model Regresi dengan Menggunakan

- Training Data

Training data adalah kumpulan data yang digunakan untuk mengajarkan model machine learning.

Training data digunakan untuk mengajarkan model prediksi yang menggunakan algoritma machine learning dengan cara mengekstrak fitur yang relevan dengan tujuan bisnis tertentu.

- Testing Data

Validation data digunakan untuk memberikan evaluasi yang tidak bias dari model fit pada training data saat menyetel hyperparameter model.

Contoh validasi yang disertakan dalam set ini akan digunakan untuk menemukan nilai optimal untuk hyperparameter dari model yang dipertimbangkan.

- Validation Data

Testing data digunakan untuk memberikan evaluasi yang tidak bias dari kecocokan model akhir pada set data pelatihan.

Testing data digunakan untuk mengevaluasi kinerja model ini dan memastikan bahwa model tersebut dapat digeneralisasikan dengan baik ke titik data baru yang tidak terlihat.

28) Forecasting analyst

Analisis deret waktu adalah cara khusus untuk menganalisis suatu rangkaian titik data yang dikumpulkan selama interval waktu tertentu. Pada waktunya analisis seri, analis mencatat titik data secara konsisten interval selama periode waktu tertentu, bukan hanya merekam titik data secara intermiten atau acak. Yang membedakan data deret waktu dengan data lainnya adalah analisis dapat menunjukkan bagaimana variabel berubah seiring waktu. Di dalam dengan kata lain, waktu adalah variabel yang krusial karena waktu menunjukkan bagaimana data menyesuaikan sepanjang titik data sebagai serta hasil akhirnya. Ini menyediakan sumber tambahan informasi dan urutan ketergantungan antara data. Analisis rangkaian waktu membantu organisasi memahami hal mendasar penyebab tren atau pola sistemik dari waktu ke waktu. Menggunakan data visualisasi, pengguna bisnis dapat melihat tren musiman dan menggali lebih dalam mengapa tren ini terjadi. Ketika organisasi menganalisis data dalam interval yang konsisten, mereka juga dapat menggunakan peramalan deret waktu untuk memprediksi kemungkinannya peristiwa masa depan. Perkiraan deret waktu adalah bagian dari analisis prediktif.

Ini dapat menunjukkan kemungkinan perubahan pada data, seperti musiman atau siklus perilaku, yang memberikan pemahaman yang lebih baik tentang variabel data dan membantu perkiraan yang lebih baik.

29) Funnel analysis

Kegunaan Funnel

1. Langkah atau steps yang user tempuh untuk membeli suatu produk
2. Share / proporsi tiap user yang lanjut ke langkah selanjutnya

Contoh funnel

Di perusahaan ecommurz, user perlu melalui tahapan berikut untuk membeli barang:

1. Masuk ke main page
2. Masuk ke product page
3. Menambahkan product ke cart
4. Masuk ke checkout page
5. Menyelesaikan order



### 30) Insight Generation

Dari funnel kita bisa lihat hanya ada 1 dari 10 user yang melanjutkan order dari halaman checkout. Apa mungkin halaman checkout kita kurang intuitif? Call to Action (CTA) yang kurang jelas? Kemudian ada 3000 pengunjung yang memasukkan item ke keranjang tapi tidak lanjut. Apakah perlu kita kirim notifikasi / email ke mereka?

Dari sini kita bisa gali lebih dalam strategi yang bisa kita lakukan untuk meningkatkan bisnis kita

### 31) Introduction tableau

Dalam 1 dekade, Tableau dinobatkan sebagai pemimpin platform dalam “Gartner Magic Quadrant for Business Intelligence and Analytics Platforms”. Hal tersebut karena Tableau memiliki beberapa keunggulan dalam memvisualisasikan data:

1. Mempersingkat waktu untuk mendapatkan insight
2. Meningkatkan akurasi untuk mendapatkan insight
3. Meningkatkan engagement

**Environment Tableau Product**

The comparison of Tableau Products

	Tableau Desktop	Tableau Reader	Tableau Public
Connectivity	Semua koneksi yang ada di Tableau	hanya membaca file workbook	Excel, file text
Distribution	Offline, Tableau Server, atau Tableau Public	Offline	Cloud (Public) Hosting service yang disediakan Tableau secara gratis
Automation	Tidak tersedia	Tidak tersedia	Tidak tersedia
Security	Kemampuan Tableau tidak lepas dari sisi server/komputer lokal	Kemampuan Tableau tidak lepas dari sisi server/komputer lokal	Terbatas, meskipun hasil yang kita buat dapat diakses oleh siapapun di web, namun kita dapat memaparkan serangan agar hasil workbook yang telah diupload tidak dapat diunduh
Best for	Siapapun yang memiliki kebutuhan untuk menghubungkan data di database, dan kebutuhan untuk membagi workbook melalui Tableau Server	Siapapun yang membutuhkan cara yang mudah untuk view dan berinteraksi dengan workbook yang sudah dibuat oleh tim lain	Instansi yang ingin membagikan insight dengan data publik, beginner yang ingin belajar menggunakan Tableau secara free, dll

## 32) Chart type

Tipe – Tipe Chart Tableau dan Fungsinya

### 1. Bar Chart (Diagram Batang)

Bar chart merupakan salah satu visualisasi yang paling umum digunakan. Chart ini biasanya digunakan untuk membandingkan data dengan cepat dari seluruh kategori yang ada.

### 2. Line Chart (Diagram Garis)

Line chart digunakan untuk melihat tren data berdasarkan runtut waktu, data dapat disajikan pertahun, quarter, bulan, bahkan hingga perhari. Hasil visualisasi yang berbentuk garis mempermudah pengguna membaca data apakah mengalami kenaikan maupun penurunan.

### 3. Pie Chart (Diagram Lingkaran)

Pie chart akan membagi lingkaran menjadi beberapa bagian sesuai dengan data. Chart ini akan lebih tepat digunakan apabila bagian-bagian tersebut tidak lebih dari 5, karena jika lebih dari itu pie chart menjadi lebih sulit dibaca.

#### 4. Maps (Geospasial)

Maps sangat mudah untuk memvisualisasikan segala jenis informasi lokasi, kode pos, negara. Jika pengguna memiliki informasi geografis yang terkait dengan data geospasial, peta merupakan cara sederhana dan menarik untuk menunjukkan bagaimana lokasi berkorelasi dengan tren di data Anda.

#### 5. Scatter Plot

Scatter Plot digunakan untuk mengetahui hubungan antara satu variabel dengan variabel lainnya yang bersifat numerik.

#### 6. Gantt Chart

Gantt Chart digunakan untuk mempermudah membaca durasi suatu aktivitas. Pada Gantt Chart setiap tanda terpisah menunjukkan durasi. Misalnya, pengguna dapat melihat metode pengiriman yang paling cepat menggunakan first class, tercermin pada pendeknya panjang chart.

#### 7. Bubble Chart

Bubble chart digunakan untuk menampilkan data dalam sekumpulan lingkaran dengan ukuran yang bervariasi. Besar kecilnya lingkaran ditentukan berdasarkan besar kecilnya nilai pada data.

#### 8. Tree Map

Tree Map digunakan untuk menampilkan visualisasi persegi panjang, besar-kecilnya persegi panjang ditentukan berdasarkan besar-kecilnya nilai pada suatu data.

### 33) Aggregate function

Aggregate functions merupakan fungsi untuk meringkas data. Di Tableau menggunakan SUM sebagai agregasi standar pada visualisasi. Untuk mengubah

function, klik kanan pada measure dan pilih Measure - lalu pilih function yang dibutuhkan.

Beberapa function yang ada di Tableau di antaranya adalah:

Fungsi	Syntax
COUNT	COUNT(expression)
COUNTD	COUNTD(expression)
MAX	MAX(expression)
MEDIAN	MEDIAN(expression)
MIN	MIN(expression)
VAR	VAR(expression)
STDEV	STDEV(expression)
SUM	SUM(expression)

### 34) Table calculations

Tableau menyediakan menu komputasi yang bisa diterapkan pada View di canvas, termasuk running total, difference, percent difference, percent of total, moving average dll. Quick Table Calculation merupakan cara cepat untuk menambahkan Table Calculation

- Keuntungan dari Table Calculation adalah:

1. Cara cepat untuk membuat advanced calculation tanpa perlu mengetahui syntaxnya seperti apa
2. Table Calculation dapat disimpan sebagai calculated field, sehingga kita dapat mempelajari syntax yang sudah dibuat oleh Tableau
3. Efficient processing karena dihitung pada subset data.

Tanda segitiga di pojok kanan pada field mengindikasikan bahwa pada field tersebut diterapkan Table Calculations.

### 35) Parameter

Parameter berguna ketika kita ingin menambahkan interaksi dan fleksibilitas pada sebuah report yang sedang kita bangun, atau melakukan skenario what-if dalam kegiatan analisis data. Parameter dapat digunakan agar User melakukan input data pada worksheet dan juga dashboard. Nilai yang diterima dapat digunakan di dalam Calculated Field atau ditampilkan sebagai reference lines pada chart yang dibuat. Tableau juga menyediakan Parameter Control (Show

Parameter Control) yang dapat ditambahkan di Worksheet dan Dashboard sehingga User dapat memilih sebuah value dari daftar (list) yang telah dibuat atau memasukkan nilai pada sebuah text field yang dapat berpengaruh pada Chart.

### 36) Dashboard dengan tableau

Dashboard pada Tableau adalah tampilan yang terdiri dari berbagai jenis visualisasi data yang digunakan untuk menyajikan informasi secara terpadu. Dashboard memungkinkan pengguna untuk menggabungkan berbagai elemen visual, seperti grafik, tabel, filter, gambar, dan teks, dalam satu tampilan yang mudah diakses. Tujuannya adalah untuk membantu pengguna memahami data dengan cara yang lebih interaktif dan informatif.

Konsep "story" dalam Tableau mengacu pada kemampuan untuk membuat narasi atau presentasi yang menghubungkan visualisasi data dalam satu aliran cerita yang koheren. Ini memungkinkan kamu untuk mengkomunikasikan temuan, insight, atau informasi penting dari visualisasi datamu kepada orang lain dengan cara yang lebih sistematis dan berurutan. Dengan menggunakan konsep "story," kamu dapat mengarahkan perhatian audiens pada berbagai aspek data yang kamu miliki dan membantu mereka memahami kisah atau pesan yang ingin kamu sampaikan.

### 37) Pentingnya visualisasi data

Visualisasi data memanfaatkan kekuatan visual otak untuk memahami informasi dengan lebih efektif. Otak kita lebih mudah mengenali pola dalam grafik daripada menganalisis angka secara terpisah. Visualisasi membantu kita menangkap tren, anomali, dan hubungan yang mungkin terlewatkan dalam data mentah. Data visual yang kuat dapat menciptakan rasa empati dan kepedulian. Lebih mudah menyampaikan pesan dan memotivasi tindakan Ketika informasi disampaikan dengan cara yang membangkitkan emosi.

### 38) Deck hasil analisis data



➤ Judul

Judul harus singkat, menarik, dan menggambarkan isi deck.

Contoh Judul: Analisis Data Konsumen

➤ Pendahuluan

Pendahuluan menjelaskan tujuan dari deck dan memberikan gambaran umum tentang temuan analisis data.

Contoh Pendahuluan :

- Tujuan: Memahami perilaku dan preferensi konsumen
- Lingkup: Konsumen di Indonesia
- Metode: Analisis data survei

➤ Temuan Utama

Temuan utama adalah poin-poin penting yang ingin disampaikan kepada stakeholder. Temuan ini harus dianalisis secara mendalam dan disajikan dengan cara yang jelas dan mudah dipahami.

Contoh Temuan Utama

- Konsumen Indonesia lebih menyukai produk lokal → Tampilkan Grafik sebagai bukti
- Konsumen Indonesia lebih memilih produk yang terjangkau → Tampilkan Grafik sebagai bukti
- Konsumen Indonesia lebih sering membeli produk secara online → Tampilkan Grafik sebagai bukti

➤ Rekomendasi

Rekomendasi adalah tindakan yang disarankan untuk diambil berdasarkan temuan utama. Rekomendasi ini harus realistis dan dapat diimplementasikan.

Contoh Rekomendasi :

- Fokus pada produk lokal
- Sesuaikan harga produk dengan daya beli konsumen

- Tingkatkan pemasaran produk secara online

➤ Kesimpulan

Kesimpulan merangkum temuan utama dan rekomendasi yang diberikan.

Contoh Kesimpulan :

Analisis data menunjukkan bahwa konsumen Indonesia memiliki beberapa preferensi yang dapat dijadikan dasar untuk pengambilan keputusan bisnis.

39) Cara mempresentasikan data

- Mengetahui audien
- Menggunakan grafik yang sederhana
- Tidak terlalu banyak detail
- Sorot poin-poin kunci untuk perhatian maksimal

B. Pembahasan Jobdesk Karya MSIB yang dikerjakan selama MSIB

Implementasi pembelajaran dengan melakukan analisis data secara berkelompok berdasarkan kasus yang telah diberikan oleh mitra, dalam hal ini yaitu Karier.mu. Karier.mu akan memberikan studi kasus nyata/ permasalahan nyata yang dapat diselesaikan oleh tiap kelompok Studi Independen Kampus Merdeka dalam bentuk proyek akhir. Tujuan proyek akhir ini adalah agar peserta mampu mengaplikasikan semua yang sudah dipelajari di dalam modul dalam menjawab pertanyaan-pertanyaan bisnis yang diberikan pada study case.

Adapun mahasiswa/i akan diminta untuk membuat dashboard visualisasi data dan *deck report* untuk meluaskan manfaat hasil program studi independen dan menambah portfolio. Setelah menyelesaikan proyek akhir, masing-masing kelompok akan mempresentasikan hasil karya pada sesi Showcase. Konsepnya akan seperti seminar publik secara daring yang menyajikan hasil pengerjaan proyek. Peserta, secara berkelompok akan menyelesaikan proyek tersebut dengan melakukan :



Gambar 4. 3 Tahapan analisis data

Untuk outputnya, setiap kelompok harus membuat 2 produk yaitu dashboard tableau dan report (deck powerpoint). Dalam hal ini, diperoleh kasus dari mitra yaitu:

#### Project 2

- Judul: Analisis Pengerjaan tes kemampuan kognitif pada platform Exam.inate
- Business Context:

Exam.inate merupakan sebuah online-assessment-platform kemampuan kognitif yang digunakan oleh Universitas Bangun Bangsa untuk menyaring kandidat mahasiswa baru. Terdapat 5 jenis tes yang dapat dipilih di dalam rangkaian ujian menyesuaikan dengan kebutuhan dari fakultas. Tingkat penyelesaian tes menjadi salah satu hal yang perlu untuk diketahui karena dengan hal tersebut Universitas dapat mengetahui berapa mahasiswa yang diterima, dan juga mengetahui biaya yang dikeluarkan ketika menggunakan platform Exam.inate.

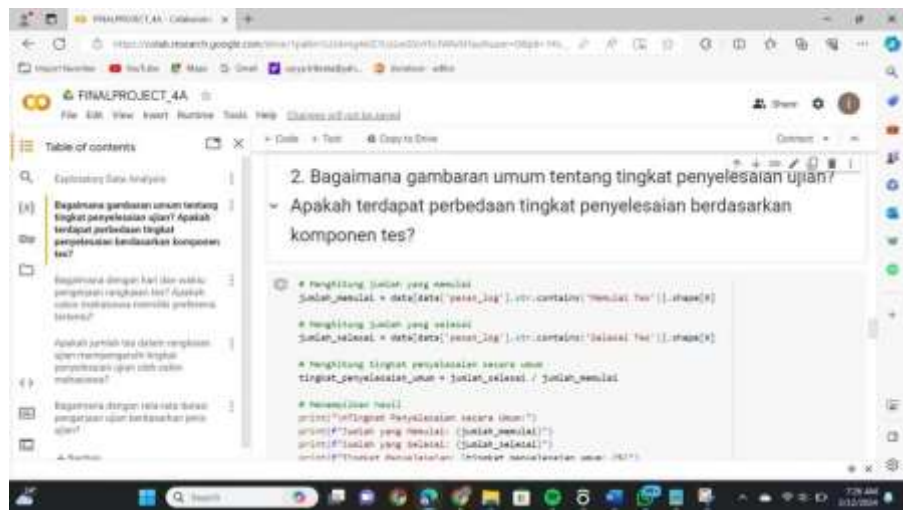
Rata-rata durasi pengerjaan tes juga menjadi metrik yang selalu diperhatikan oleh Exam.inate agar waktu yang diberikan untuk menyelesaikan tes bisa disesuaikan dengan tingkat kesulitan dari tes.

- Problem Questions
  - 1) Bagaimana dengan kondisi datanya? Apakah terdapat duplikasi dan missing value?
  - 2) Bagaimana gambaran umum tentang tingkat penyelesaian ujian? Apakah terdapat perbedaan tingkat penyelesaian berdasarkan komponen tes?

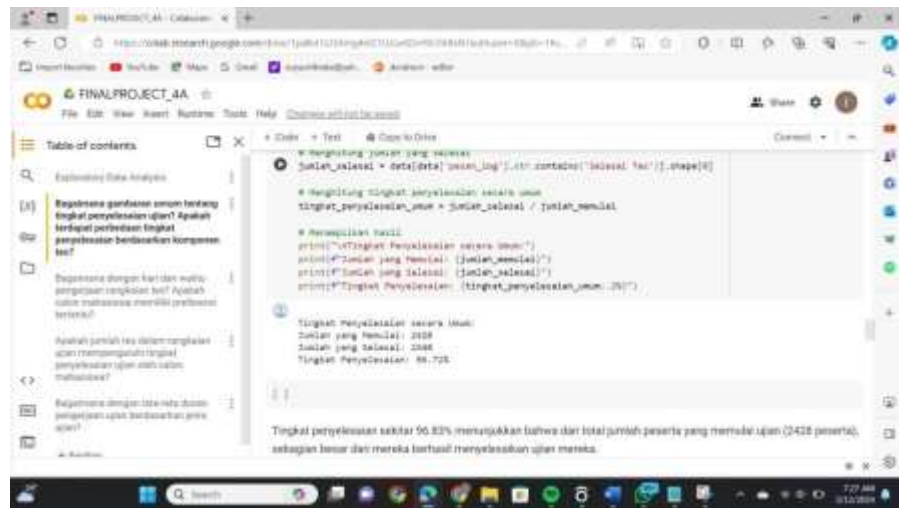
- 3) Bagaimana dengan hari dan waktu pengerjaan rangkaian tes? Apakah calon mahasiswa memiliki preferensi tertentu?
- 4) Apakah jumlah tes dalam rangkaian ujian mempengaruhi tingkat penyelesaian ujian oleh calon mahasiswa?
- 5) Apa saja rekomendasi terkait jumlah rangkaian tes, jenis tes yang digunakan dan periode waktu ujian yang dapat diberikan kepada universitas untuk menaikkan tingkat penyelesaian ujian?

Berdasarkan kasus diatas maka diperoleh hasil berupa:

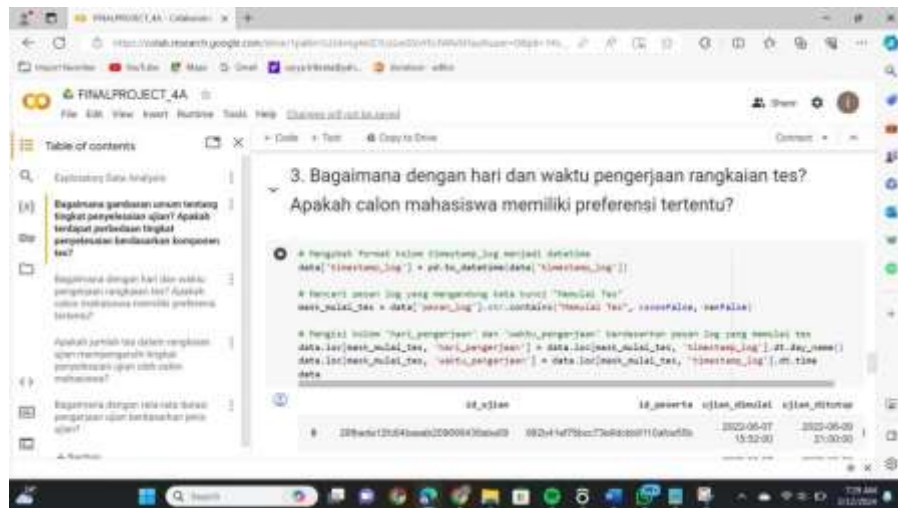
➤ Google Collabs



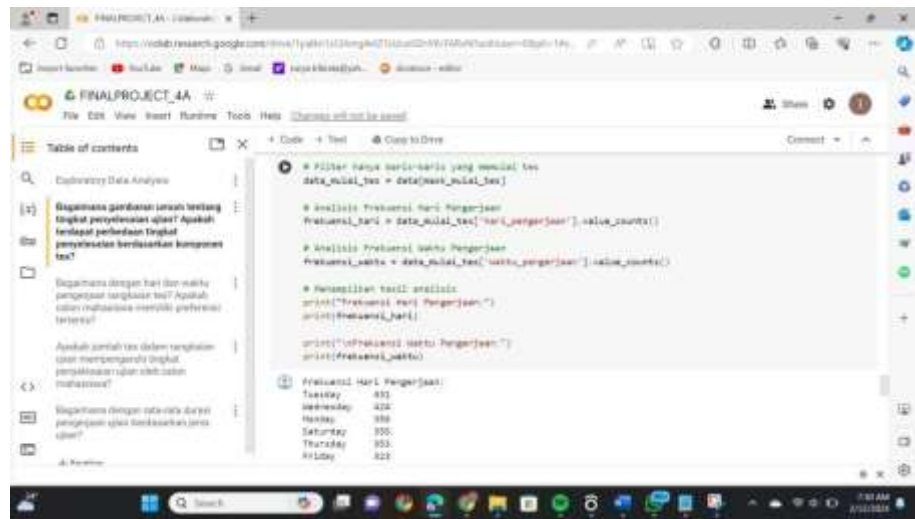
Gambar 4. 4 Google Collabs Pembahasan No 2



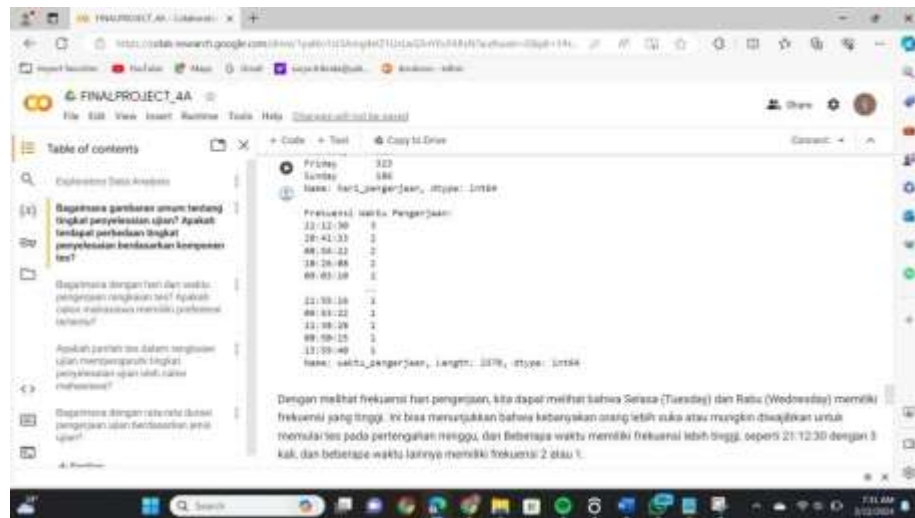
Gambar 4. 5 Google Collabs Pembahasan No 2



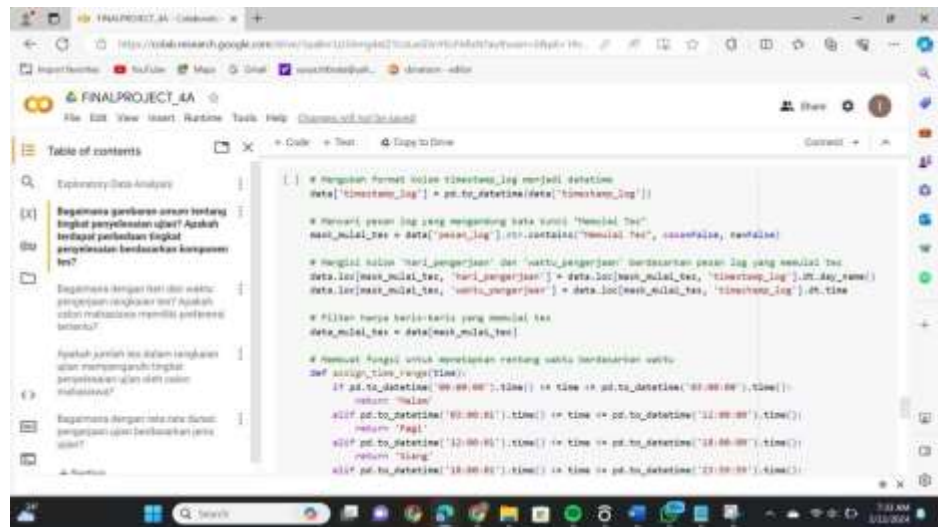
Gambar 4. 6 Google Collabs Pembahasan No 3



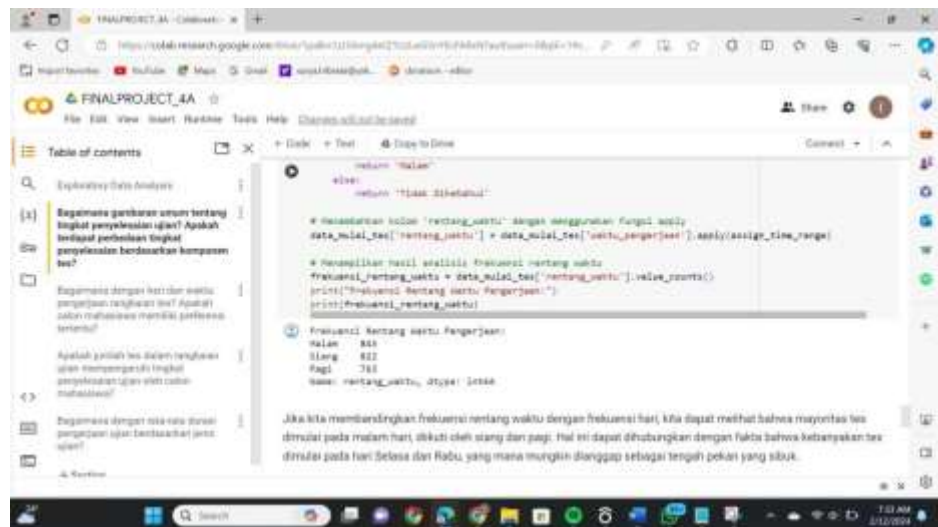
Gambar 4. 7 Google Collabs Pembahasan No 3



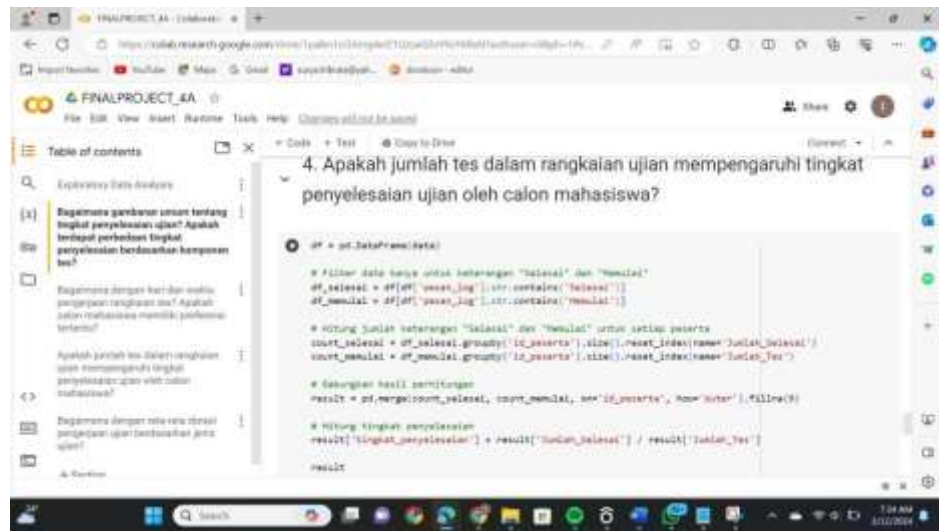
Gambar 4. 8 Google Collabs Pembahasan No 3



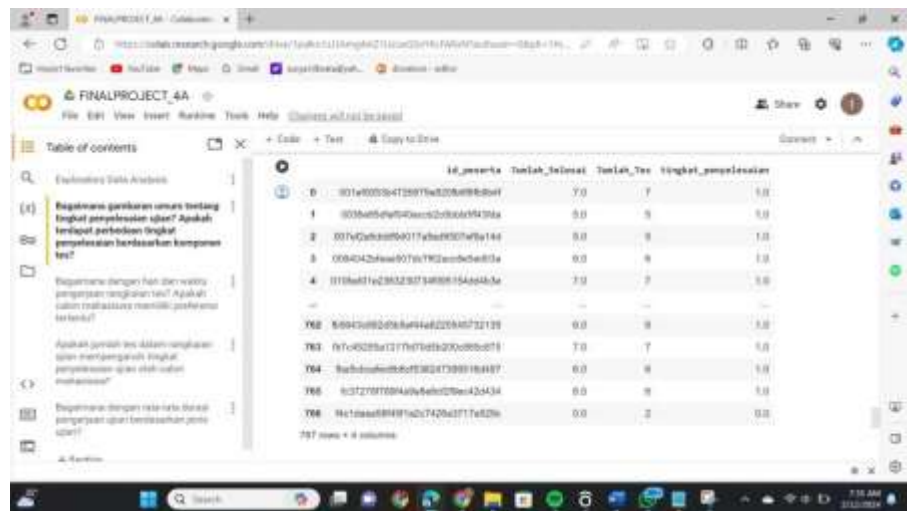
Gambar 4. 9 Google Collabs Pembahasan No 3



Gambar 4. 10 Google Collabs Pembahasan No 3

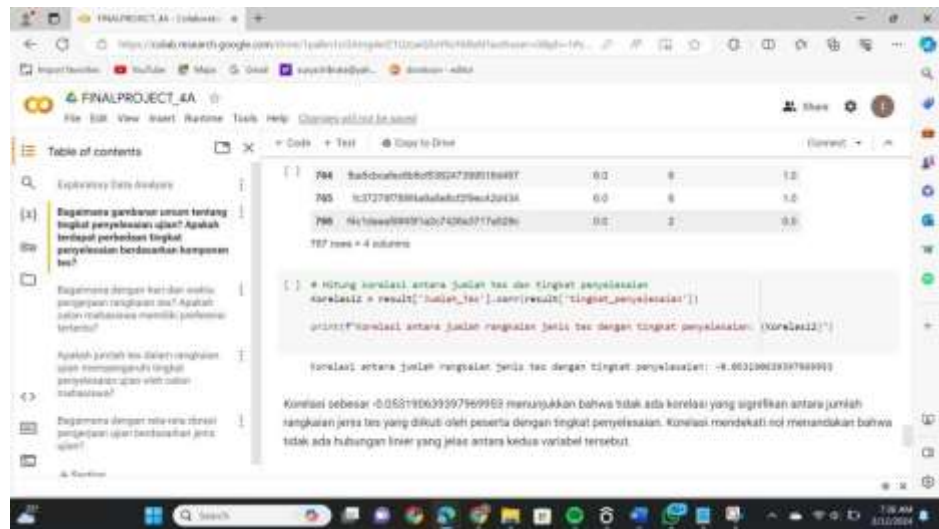


Gambar 4. 11 Google Collabs Pembahasan No 4

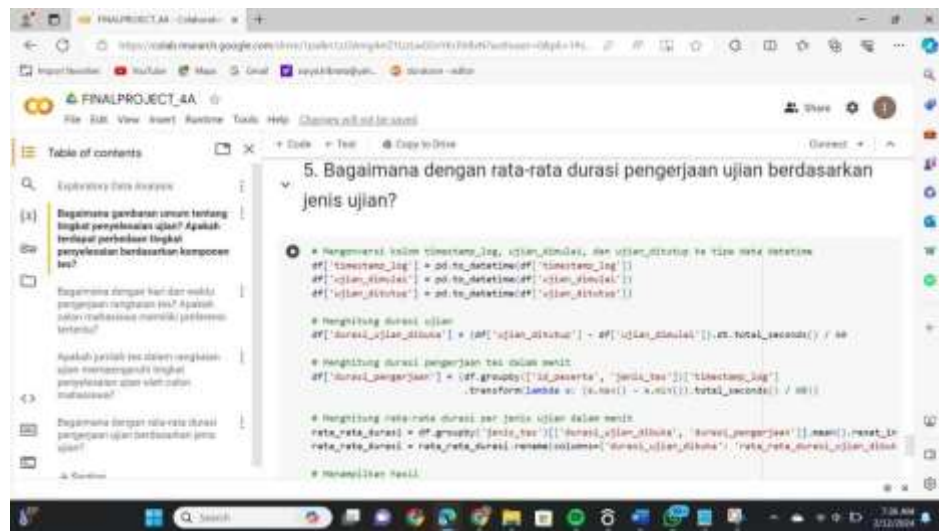


Gambar 4. 12 Google Collabs Pembahasan No 4

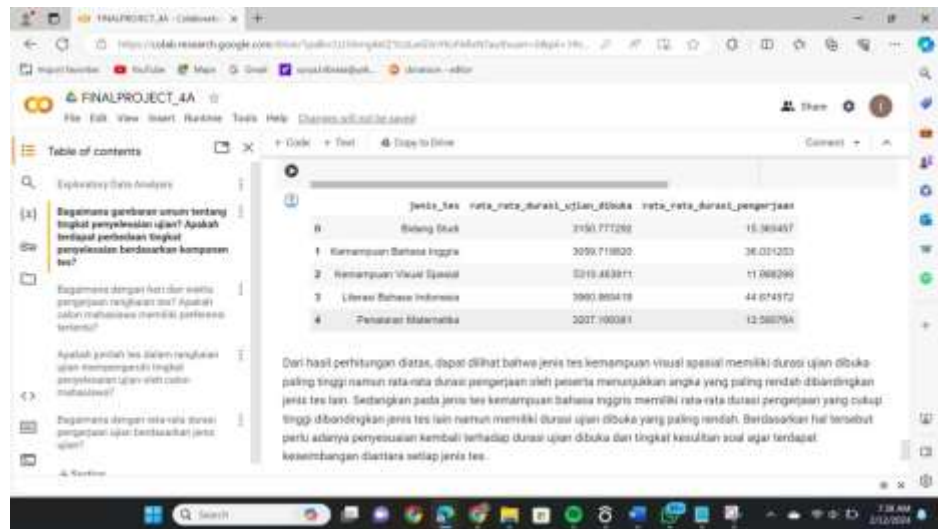




Gambar 4. 13 Google Collabs Pembahasan No 4



Gambar 4. 14 Google Collabs Pembahasan No 5

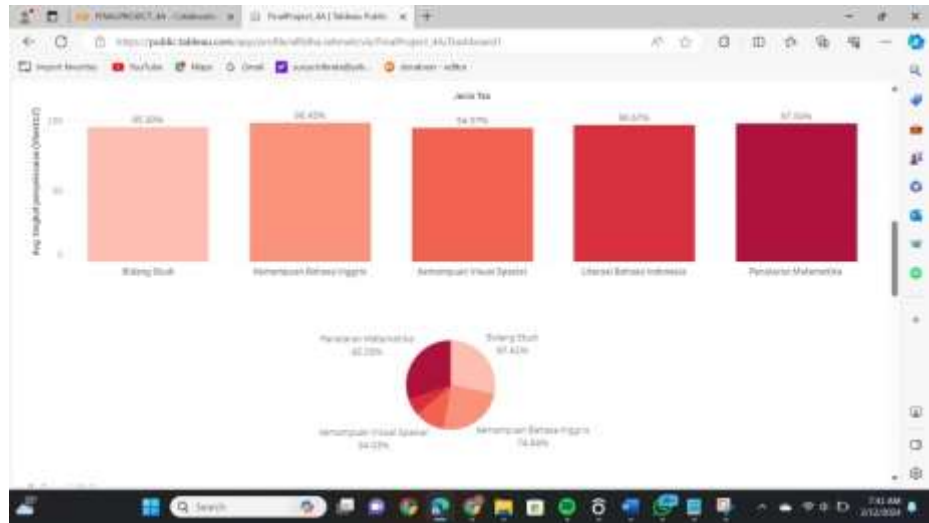


Gambar 4. 15 Google Collabs Pembahasan No 5

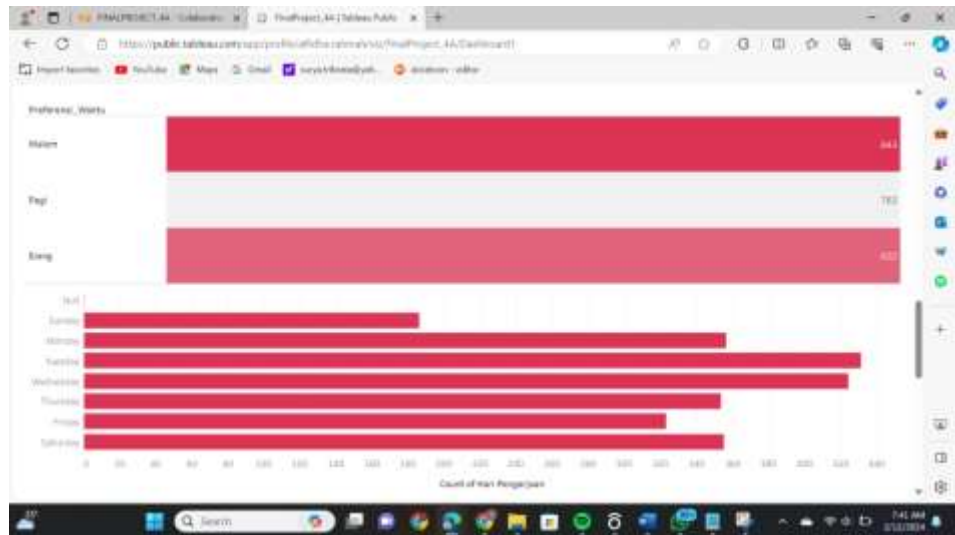
➤ Visualisasi dengan tableau



Gambar 4. 16 Visualisasi Jenis Tes



Gambar 4. 17 Visualisasi Jenis Tes



Gambar 4. 18 Visualisasi Hari Pengerjaan



Gambar 4. 19 Visualisasi Hari pengerjaan

➤ Deck Powerpoint



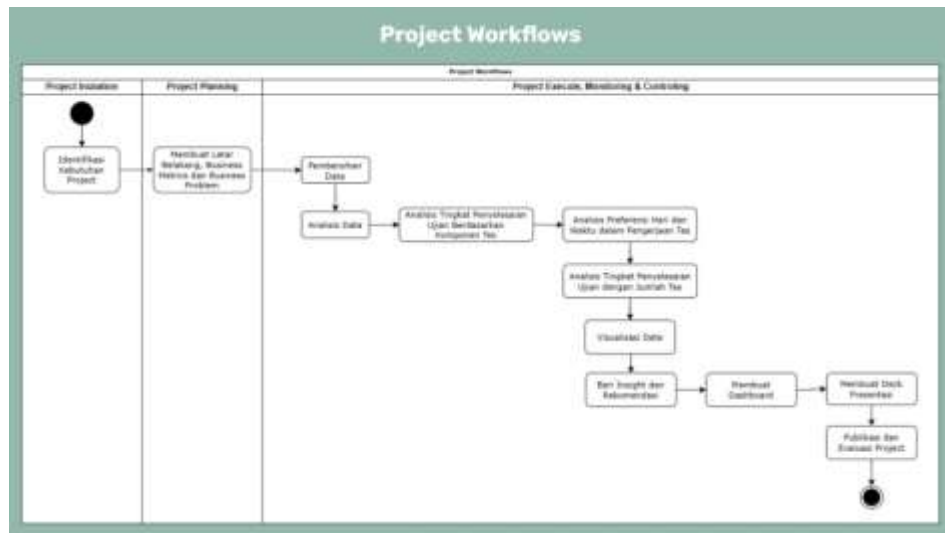
Gambar 4. 20 Deck Powerpoint Latar Belakang



Gambar 4. 21 Deck Powerpoint Latar Belakang



Gambar 4. 22 Deck Powerpoint Bussines Problem



Gambar 4. 23 Deck Powerpoint Project Workflows



Gambar 4. 24 Deck Powerpoint Data Cleaning



Gambar 4. 25 Deck Powerpoint Pembahasan No 1



Gambar 4. 26 Deck Powerpoint Pembahasan No 2



Gambar 4. 27 Deck Powerpoint Pembahasan No 2

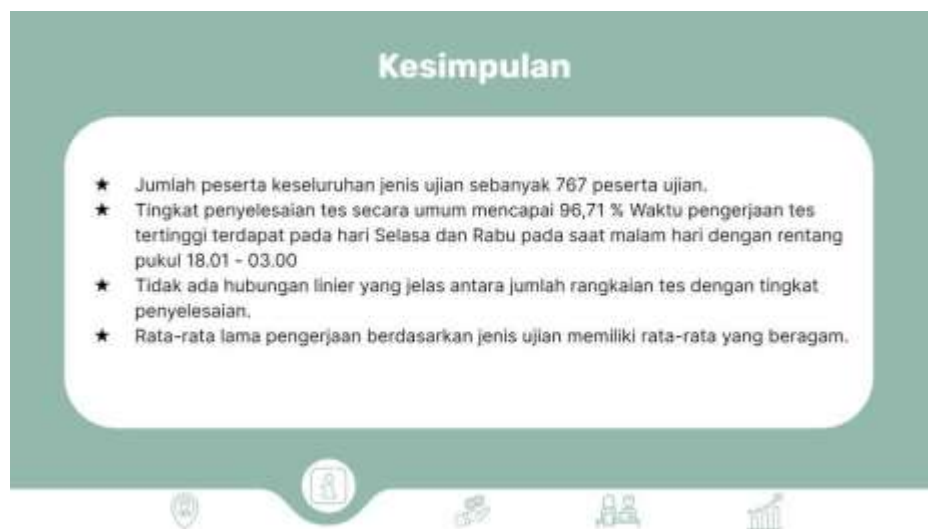


Gambar 4. 28 Deck Powerpoint Pembahasan No 3





Gambar 4. 29 Deck Powerpoint Pembahasan No 4



Gambar 4. 30 Deck Powerpoint Kesimpulan



Gambar 4. 31Deck Powerpoint Rekomendasi

## **BAB V**

### **PENUTUP**

#### **A. Kesimpulan**

Berdasarkan program studi independent yang telah dilakukan selama ini, dapat diambil kesimpulan sebagai berikut:

1. Program studi independen melatih untuk belajar dan mengembangkan diri melalui aktivitas di luar kelas perkuliahan, dengan menguasai kompetensi spesifik dan praktis serta mendapatkan isu nyata yang ada pada dunia industri.
2. Program studi independent memberikan pengalaman dan pemahaman materi yang lebih dalam yang mana sebelumnya pada saat diperkuliahan belum diperoleh.
3. Materi yang disampaikan sangat baik, serta expert yang kompeten dibidangnya.
4. Kemampuan yang cukup terlatih karena setiap modul dilengkapi dengan pretest, quiz, praktik mandiri serta posttest. Hasil dari studi independent yaitu proyek akhir yang dihasilkan dari kerja tim, mulai dari mengidentifikasi kebutuhan proyek hingga dapat memberikan insight dan rekomendasi yang tepat. Selanjutnya, hasil tersebut dipresentasikan dihadapan tim kariermu dan juri.

#### **B. Saran**

Saran mengenai pelaksanaan program MSIB Data Analys di PT Semesta Integrasi Digital adalah:

1. Disarankan agar video pembelajaran pada LMS dapat didownload agar materi dapat ditonton kapanpun.
2. Disarankan untuk latihan soal seperti quiz, pretest dan posttest diakhir pembelajaran terdapat pembahasan atau kunci jawaban, sehingga dapat digunakan sebagai referensi belajar untuk kedepannya.

3. Disarankan untuk ringkasan materi, catatan tersebut dapat dilakukan copy paste, agar dapat dipindahkan dengan mudah ke catatan pribadi dan dapat digunakan untuk belajar kapanpun tanpa harus membuka LMS.

## Lampiran

i. Brosur

Tidak ada

ii. Sertifikat

<https://drive.google.com/drive/folders/1DDBsGafGfT3NLCvxbyPgntml0s03CIPa?usp=sharing>

iii. Logbook

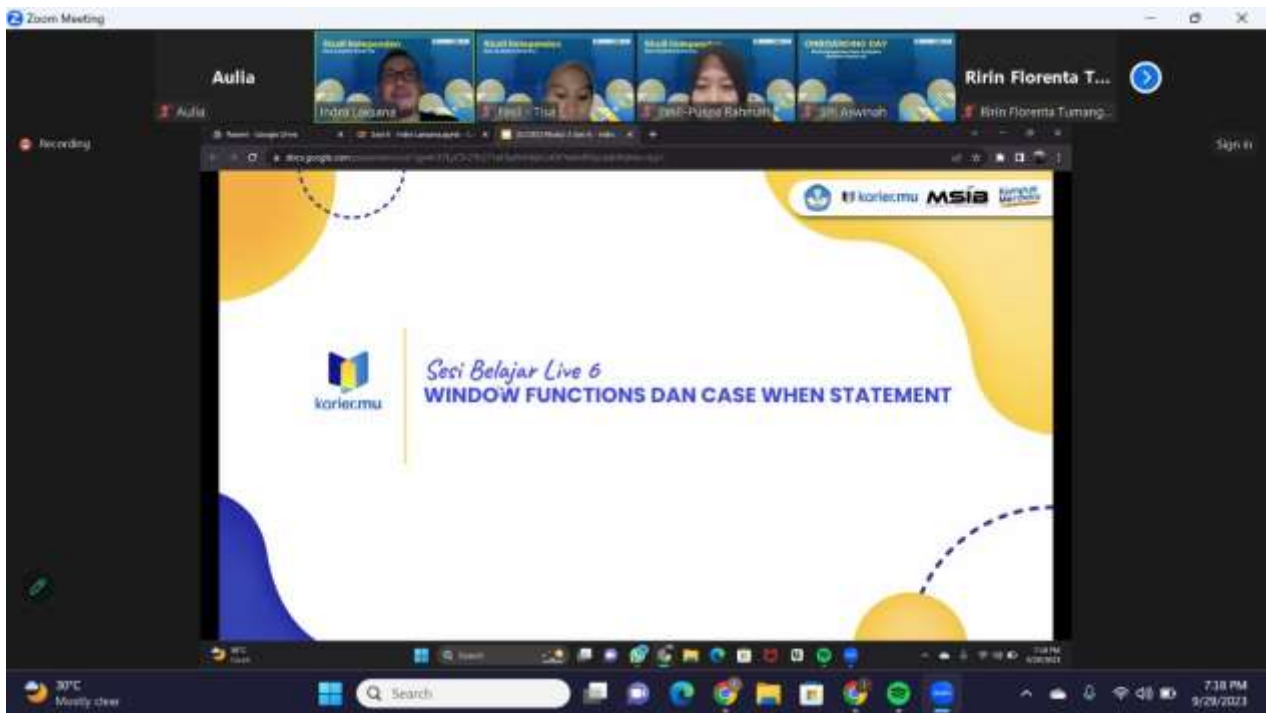
<https://drive.google.com/file/d/1LpA4ynmY8FH6MOsaqOfepiRHpY3jiYY3/view?usp=sharing>

iv. Dokumentasi Kegiatan

The image is a screenshot of a Zoom meeting. The main window displays a presentation slide with the title "Tahapan Cleaning Data" in blue. Below the title is a list of six steps:

1. Mendeteksi kesalahan data
2. Melakukan cek duplikasi data
3. Memperbaiki kesalahan struktur data
4. Filter outlier yang tidak diinginkan
5. Tangani data yang hilang
6. Validasi data dan lakukan *quality assurance*

To the right of the list is an illustration of a funnel with data points entering and exiting. The Zoom interface includes a top bar with participant names (Egi Ribeka, etc.) and a "Recording" indicator. On the right side, a chat window is open, showing a message from "Yael Punga Rahmah" with a link to a Google Drive folder: <https://drive.google.com/drive/folders/1DDBsGafGfT3NLCvxbyPgntml0s03CIPa?usp=sharing>. The Windows taskbar at the bottom shows the date as 8/28/2023 and the time as 3:30 PM.



- v. Link Dashboard  
[http://bit.ly/Dashboard4A\\_Project2](http://bit.ly/Dashboard4A_Project2)
- vi. Link Report Deck  
<https://bit.ly/3U0fZ3J>
- vii. Link Google Collabs  
[https://bit.ly/GoogleColab4A\\_Project2](https://bit.ly/GoogleColab4A_Project2)