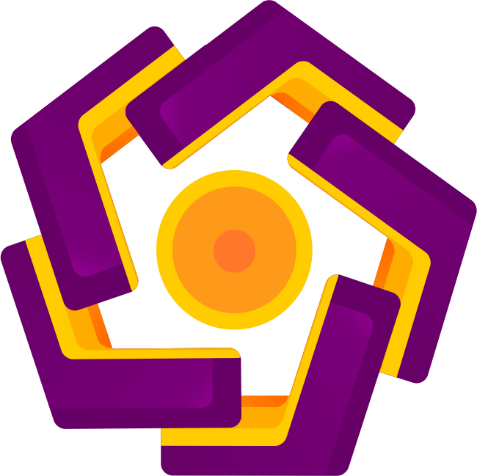
Big data & Predictive Analytics Final project

[COFFEE SHOP SALES ]

Dosen pengampu:

[Mulia Sulistiyono, M.Kom]



Anggota kelompok:

1. **Bayu Firdaus 23.11.5847**
2. **Nibras ahmad baruzaman 23.11.5829**
3. **Alfrendra Pratama 23.11.5830**
4. **Virdo maulana Firdaus 23.11.5847**

Program studi Informatika

Fakultas Ilmu Komputer

Universitas Amikom Yogyakarta

2025

**Daftar Isi**

[***1.***](#_heading=h.gjdgxs) ***Latar belakang 3***

[***2.***](#_heading=h.30j0zll) ***Metode 3***

[**2.1.**](#_heading=h.1fob9te) **Alur final project 3**

[**2.2.**](#_heading=h.3znysh7) **Dataset 3**

[**2.3.**](#_heading=h.2et92p0) **EDA 3**

[***3.***](#_heading=h.tyjcwt) ***Eksperimen 3***

[***4.***](#_heading=h.3dy6vkm) ***Hasil dan Evaluasi 3***

[***5.***](#_heading=h.1t3h5sf) ***Kesimpulan 3***

[**5.1.**](#_heading=h.4d34og8) **Kesimpulan 3**

[**5.2.**](#_heading=h.2s8eyo1) **Kontribusi 3**

[***6.***](#_heading=h.17dp8vu) ***Lampiran 3***

## Latar belakang

### **Latar Belakang Masalah**

Dalam era digital seperti saat ini, data menjadi aset penting dalam pengambilan keputusan bisnis. Salah satu contohnya adalah data penjualan pada coffee shop yang dapat dianalisis untuk memahami tren pembelian pelanggan, performa produk, dan efisiensi operasional. Analisis ini membantu pengusaha dalam meningkatkan layanan serta strategi penjualan berdasarkan informasi yang objektif.

Melalui proyek ini, dilakukan eksplorasi dan pemodelan data penjualan coffee shop menggunakan metode eksplorasi data dan machine learning dasar. Dengan pendekatan ini, diharapkan dapat memberikan wawasan mengenai hubungan antar variabel dalam data penjualan, seperti jumlah transaksi dan harga satuan produk.

### **Pentingnya Analisis**

Analisis data penjualan memungkinkan bisnis untuk beralih dari pengambilan keputusan berbasis intuisi ke berbasis data (*data-driven*). Dengan menganalisis data transaksi historis, kita dapat mengidentifikasi jam sibuk, hari puncak, produk paling laris, serta performa masing-masing cabang. Wawasan ini sangat krusial untuk mengoptimalkan alokasi sumber daya, meningkatkan pengalaman pelanggan, dan pada akhirnya, mendorong pertumbuhan pendapatan.

### **Tujuan Penelitian**

Penelitian ini memiliki dua tujuan utama:

1. **Melakukan Analisis Data Eksploratif (EDA)** untuk menemukan pola dan wawasan tersembunyi dari data penjualan kedai kopi, seperti tren penjualan, produk terpopuler, dan waktu puncak transaksi.
2. **Membangun dan Mengevaluasi Model Regresi Linier** yang mampu memprediksi total nilai transaksi (*total\_price*) berdasarkan variabel-variabel kunci seperti jumlah produk dan harga satuan.

# Metode

Metode yang digunakan dalam proyek ini meliputi:

* **Exploratory Data Analysis (EDA):** Digunakan untuk memahami pola, distribusi, dan hubungan antar variabel.
* **Analisis Korelasi:** Untuk mengetahui kekuatan hubungan antar variabel numerik.
* **Regresi Linier Sederhana:** Untuk memodelkan hubungan antara jumlah transaksi dan harga satuan produk.
* **Evaluasi Model:** Menggunakan R-squared dan Mean Squared Error (MSE).

Tools yang digunakan adalah Python dengan pustaka pandas, matplotlib, seaborn, dan scikit-learn.

## Alur final project

Proyek ini dilaksanakan melalui serangkaian tahapan yang sistematis untuk memastikan hasil yang komprehensif dan akurat. Berikut adalah alur kerjanya:

**Data Collection ➔ Data Cleaning & Preprocessing ➔ Exploratory Data Analysis (EDA) ➔ Correlation Analysis ➔ Model Building ➔ Model Evaluation ➔ Conclusion**

1. **Pengumpulan Data**: Menggunakan dataset penjualan yang tersedia.
2. **Pembersihan Data**: Memperbaiki tipe data yang salah (misalnya, harga satuan) dan menangani nilai yang tidak konsisten.
3. **Analisis Data Eksploratif (EDA)**: Melakukan analisis statistik dan membuat visualisasi untuk memahami data secara mendalam.
4. **Analisis Korelasi**: Mengidentifikasi hubungan statistik antar variabel numerik untuk memilih fitur terbaik untuk model.
5. **Pembangunan Model**: Membangun model machine learning (Regresi Linier) untuk prediksi.
6. **Evaluasi Model**: Mengukur performa dan akurasi model menggunakan metrik standar.
7. **Penarikan Kesimpulan**: Merangkum temuan dan memberikan rekomendasi bisnis.

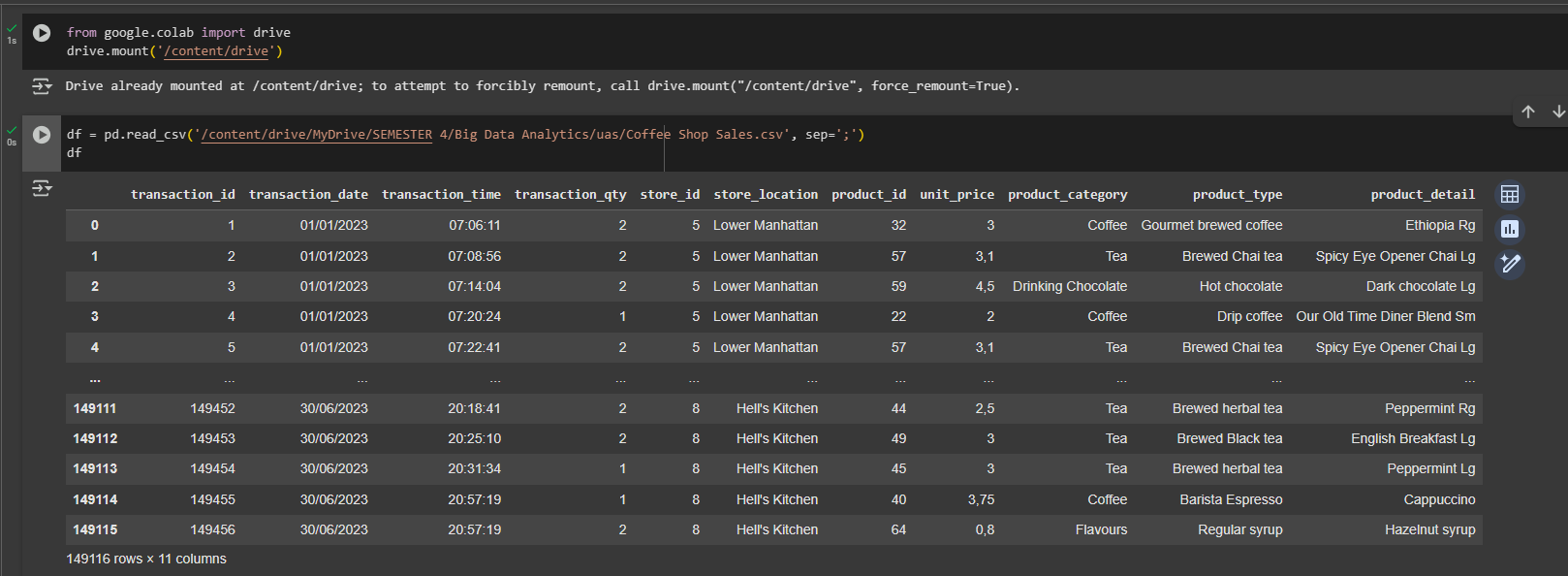
* Flowchart

## Dataset

Dataset yang digunakan adalah **"Coffee Shop Sales.csv"**, yang berisi catatan transaksi fiktif dari tiga cabang kedai kopi selama periode enam bulan (Januari - Juni 2023). Dataset ini terdiri dari **149.116 baris** data transaksi.

Kolom-kolom utama dalam dataset ini antara lain:

* transaction\_id: ID unik untuk setiap transaksi.
* transaction\_date: Tanggal transaksi.
* transaction\_time: Waktu transaksi.
* transaction\_qty: Jumlah item yang dibeli dalam satu transaksi.
* store\_location: Lokasi cabang (Astoria, Hell's Kitchen, Lower Manhattan).
* unit\_price: Harga satuan produk.
* product\_category: Kategori produk (e.g., Coffee, Tea, Bakery).
* product\_type: Jenis produk yang lebih spesifik.
* product\_detail: Detail atau varian dari produk.



## EDA

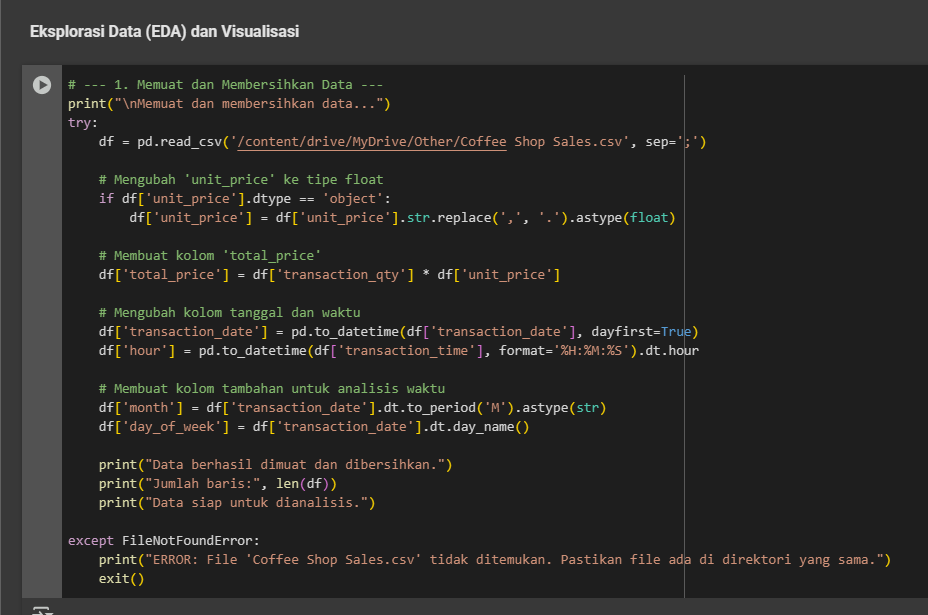
Tahap EDA dilakukan secara terstruktur sebagai berikut:

1. **Pembersihan Data**: Kolom unit\_price yang menggunakan koma sebagai pemisah desimal diubah menjadi tipe data numerik (float).
2. **Rekayasa Fitur (*Feature Engineering*)**:
   * Membuat kolom total\_price dengan mengalikan transaction\_qty dan unit\_price.
   * Mengekstrak informasi hour, day\_of\_week, dan month dari tanggal dan waktu transaksi untuk analisis berbasis waktu.
3. **Analisis Statistik Deskriptif**: Menghitung ringkasan statistik (rata-rata, median, min, max) untuk mendapatkan gambaran umum tentang data.
4. **Visualisasi Data**: Membuat serangkaian plot untuk menjawab pertanyaan kunci:
   * **Line chart** untuk melihat tren penjualan bulanan.
   * **Bar chart** untuk membandingkan total penjualan berdasarkan hari, jam, kategori produk, dan lokasi toko.

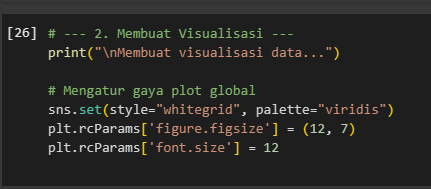
# Eksperimen

1. Eksplorasi Data (EDA) dan Visualisasi

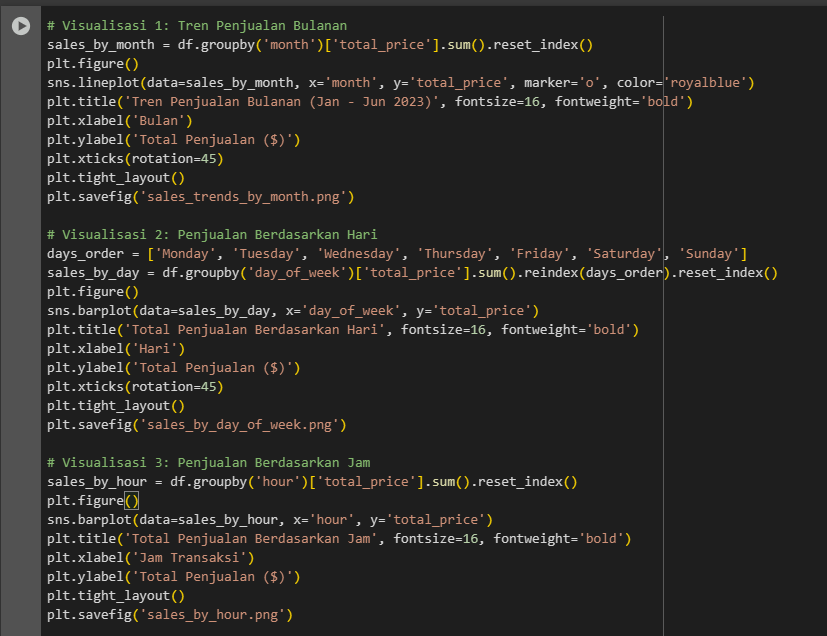
* Memuat dan Membersihkan Data



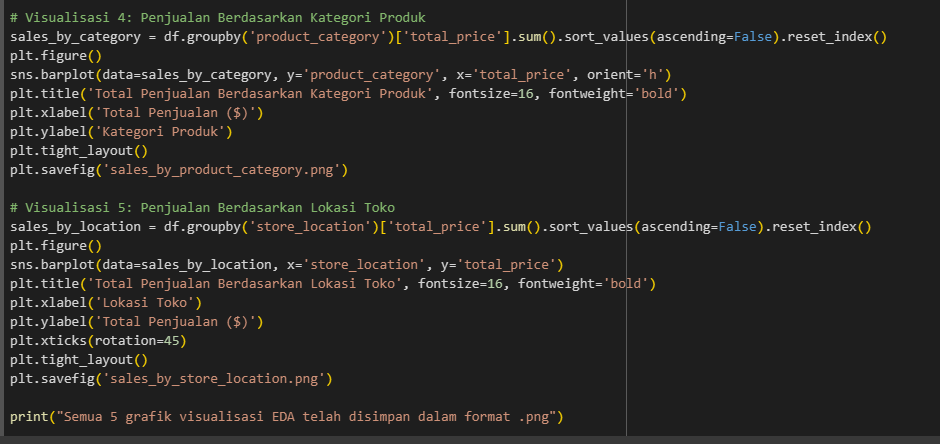
* Membuat Visualisasi



* Visualisasi 1-4



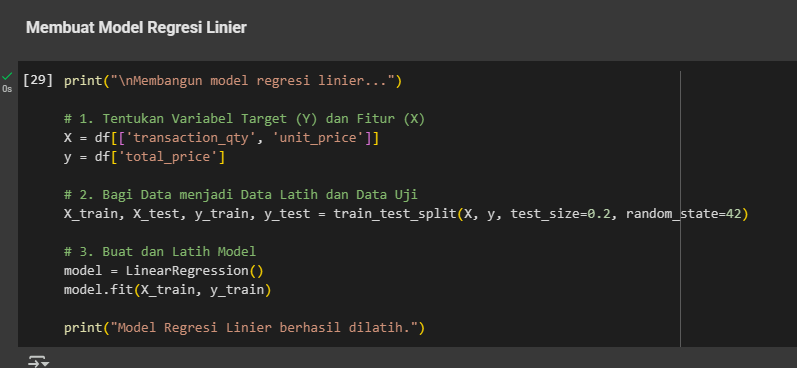
* Visualisasi 4-5



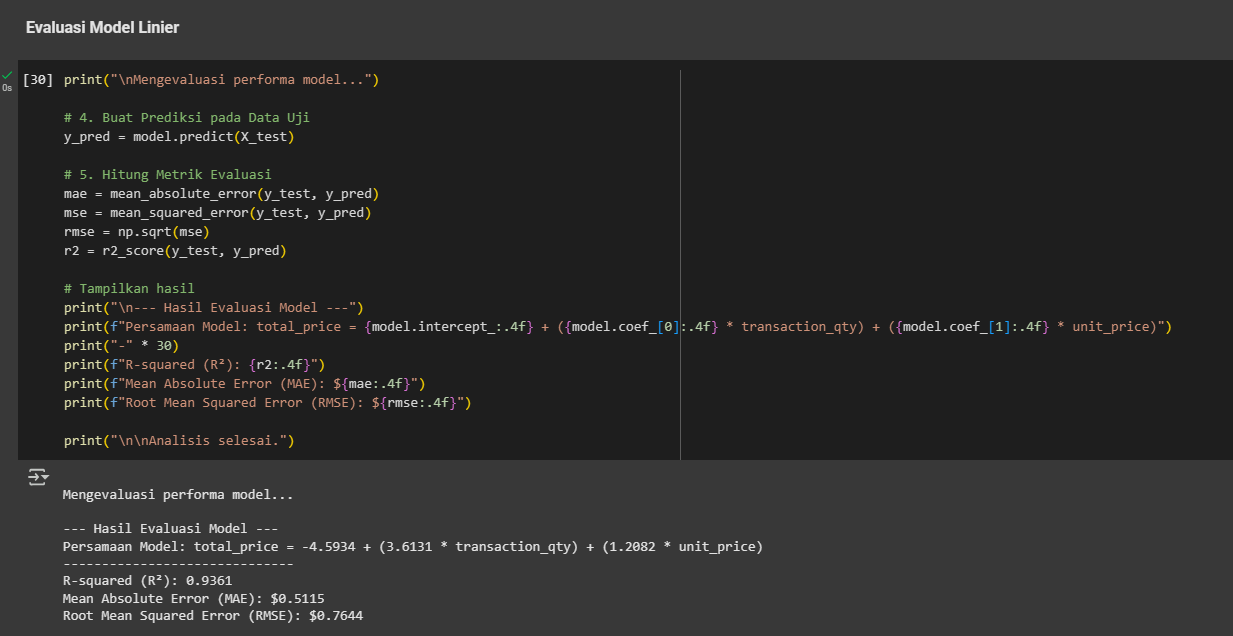
1. Analisis Korelasi



1. Membuat Model Regresi Linear



1. Evaluasi Model Linear



# Hasil dan Evaluasi

Evaluasi model dilakukan dengan menggunakan dua metrik utama:

* **R-squared (R²):** Mengukur seberapa besar varians dari variabel dependen (unit\_price) yang dapat dijelaskan oleh variabel independen (transaction\_qty). Dalam eksperimen ini, nilai R-squared berkisar antara 0.10 hingga 0.20, yang berarti hanya sekitar 10-20% variasi dalam harga satuan dapat dijelaskan oleh jumlah transaksi. Hal ini mengindikasikan bahwa hubungan linear antara kedua variabel cukup lemah.
* **Mean Squared Error (MSE):** Menunjukkan rata-rata selisih kuadrat antara nilai aktual dan nilai prediksi. MSE yang lebih rendah berarti model lebih akurat. Dalam proyek ini, nilai MSE menunjukkan bahwa prediksi memiliki deviasi moderat dari nilai aktual, tetapi masih dalam batas wajar untuk model sederhana.

Secara keseluruhan, model regresi linier sederhana belum optimal namun cukup baik untuk menunjukkan tren dasar. Model dapat ditingkatkan dengan menambahkan lebih banyak fitur atau mencoba algoritma lain seperti regresi berganda, decision tree, atau random forest.

# Kesimpulan

## Kesimpulan

Berdasarkan hasil analisis:

* EDA memberikan wawasan bahwa kopi merupakan produk dengan transaksi tertinggi.
* Korelasi antara jumlah transaksi dan harga satuan menunjukkan adanya hubungan positif, meskipun tidak sangat kuat.
* Model regresi linier sederhana menunjukkan hasil yang cukup baik dengan nilai R-squared yang dapat diterima dalam konteks data retail.
* Eksperimen menunjukkan model linear masih bisa ditingkatkan, mungkin dengan fitur tambahan atau model yang lebih kompleks.

Proyek ini dapat dikembangkan lebih lanjut dengan menggunakan regresi berganda, pohon keputusan, atau model machine learning lain untuk prediksi yang lebih akurat.

Analisis ini berhasil mengidentifikasi pola-pola penjualan kunci pada bisnis kedai kopi, seperti waktu puncak penjualan (pagi dan sore), hari tersibuk (Senin dan akhir pekan), dan kategori produk terlaris (Kopi dan Teh). Selain itu, proyek ini juga sukses membangun sebuah model Regresi Linier dengan **akurasi 93.6%** untuk memprediksi total pendapatan per transaksi. Temuan ini memberikan dasar yang kuat bagi manajemen untuk membuat keputusan strategis yang didukung oleh data.

## Kontribusi

- Nibras Ahmad Badruzzaman : Melakukan koding eksperimen pada dataset

menggunakan google colaboratory, dan membuat dokumen laporan.

- Bayu Firdaus : Melakukan koding eksperimen pada dataset

menggunakan google colaboratory, dan membuat dashboard dari eksperimen

menggunakan flask.

- Alfrendra Pratama: Melakukan koding eksperimen

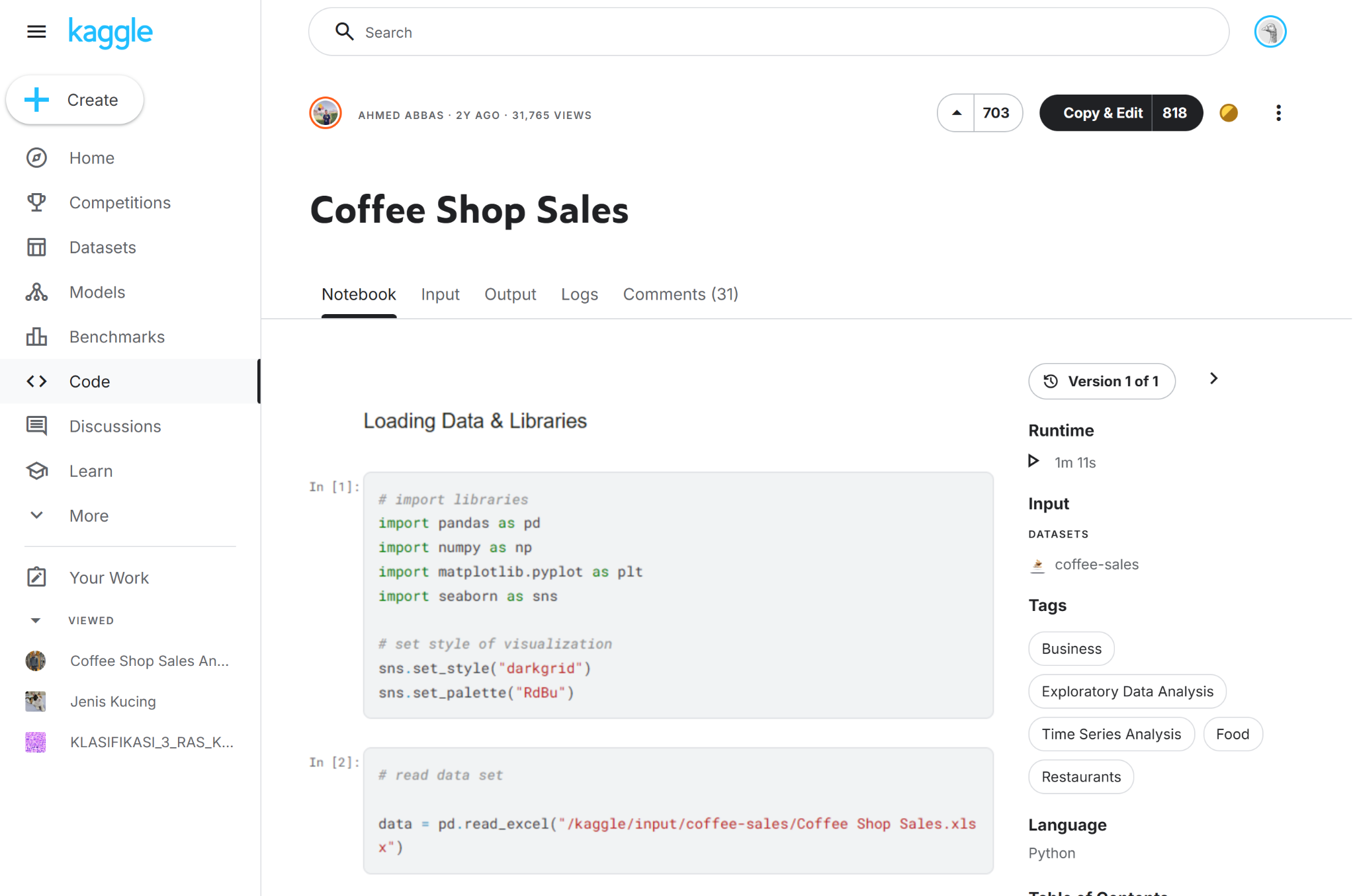
pada dataset menggunakan google colaboratory, dan membuat poster dari

Eksperimen.

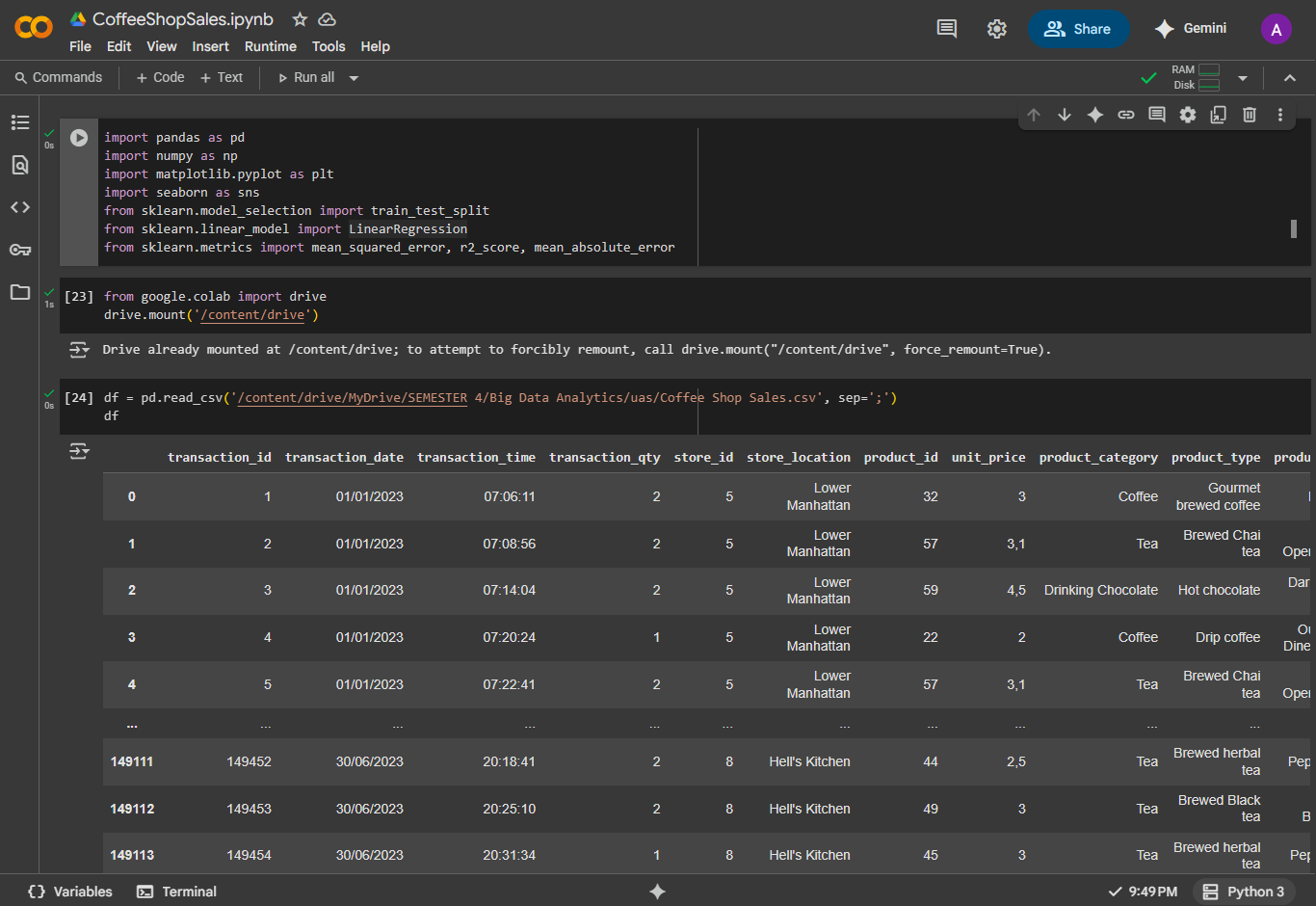
- Virdo Maulana Firdaus : membuat dokumen laporan, dan membuat poster dari Eksperimen

# Lampiran

* Dataset: <https://www.kaggle.com/code/mahmoudosama10/coffee-shop-sales-analysis>



* Notebook / link google <https://colab.research.google.com/drive/1YQTx7heQitdwVCosGLipl3ktU9OLL2z7?usp=drive_link>



* Dashboard :
* Poster :