



PORTOFOLIO

BY: RYNO PAHLEVI AL GHIFARI

RYNOPAHLEVIA@GMAIL.COM

Data Visualization



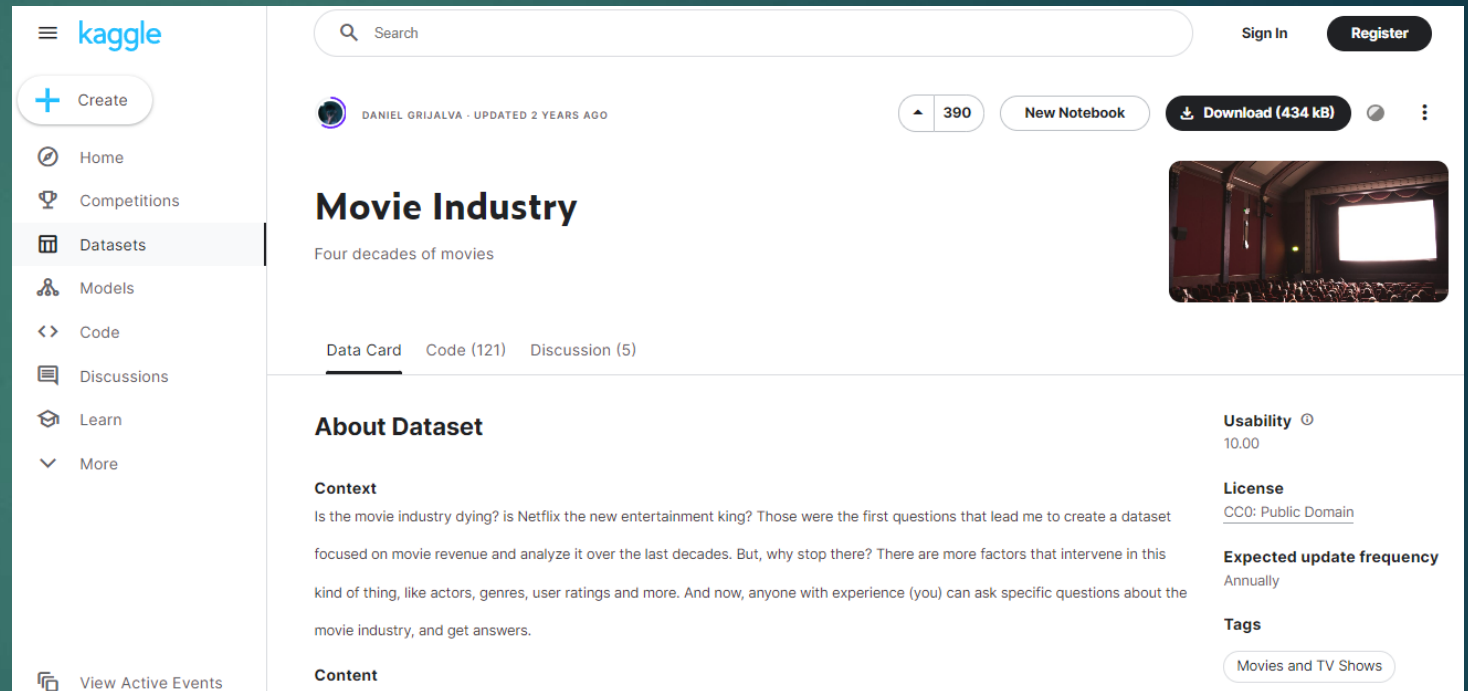
BY: RYNO PAHLEVI AL GHIFARI

RYNOPAHLEVIA@GMAIL.COM

Movie Industry Dataset

- All the data utilized in this analysis are sourced from [Kaggle](https://www.kaggle.com/danielgrijalva/movie-industry).

Within the comprehensive dataset, which spans the years from 1986 to 2016, a total of 6820 movies have been included for examination and study.



The screenshot displays the Kaggle interface for the 'Movie Industry' dataset. On the left is a navigation sidebar with links to Home, Competitions, Datasets (highlighted), Models, Code, Discussions, Learn, and More. The main content area shows the dataset title 'Movie Industry' by Daniel Grijalva, updated 2 years ago, with 390 views and options to create a new notebook or download the 434 kB dataset. Below the title is a 'Data Card' tab. The 'About Dataset' section includes a 'Context' paragraph about the dataset's focus on movie revenue and industry factors. On the right, a sidebar provides additional details: 'Usability' (10.00), 'License' (CC0: Public Domain), 'Expected update frequency' (Annually), and 'Tags' (Movies and TV Shows).

Kaggle

Search

Sign In Register

390 New Notebook Download (434 kB)

Movie Industry

Four decades of movies

Data Card Code (121) Discussion (5)

About Dataset

Context

Is the movie industry dying? Is Netflix the new entertainment king? Those were the first questions that lead me to create a dataset focused on movie revenue and analyze it over the last decades. But, why stop there? There are more factors that intervene in this kind of thing, like actors, genres, user ratings and more. And now, anyone with experience (you) can ask specific questions about the movie industry, and get answers.

Content

Usability 10.00

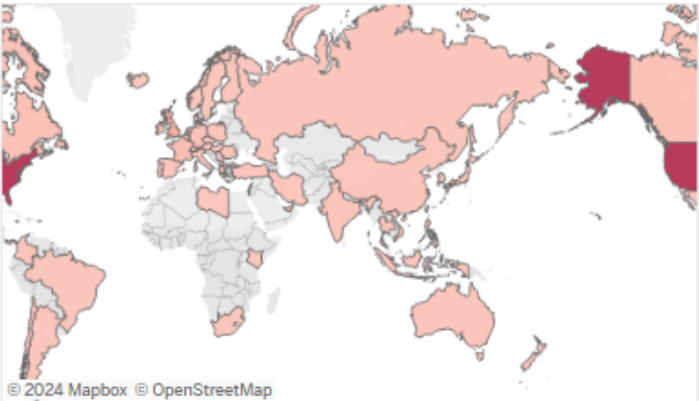
License CC0: Public Domain

Expected update frequency Annually

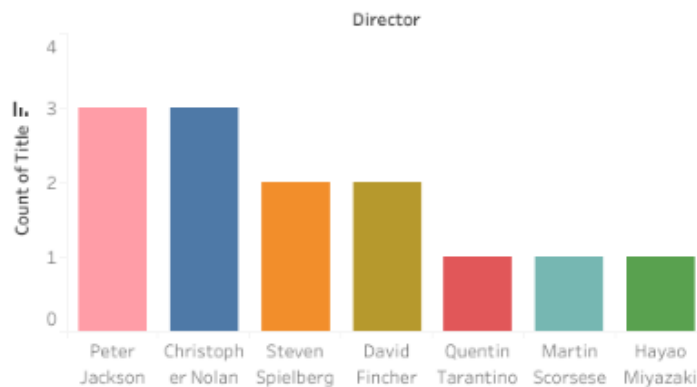
Tags Movies and TV Shows

Movie Dashboard

Country With Most Movies



Director With The Most Movies Ratings Higher Than 8.5



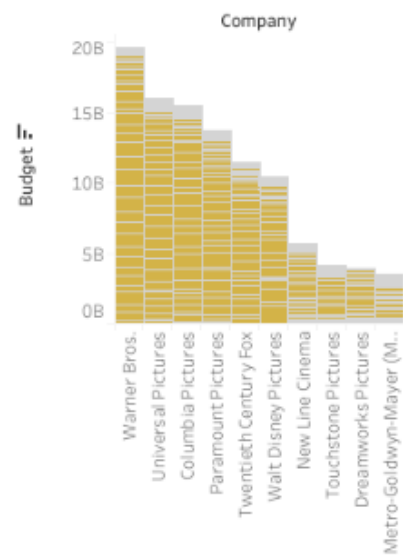
Movie Title List

Title	Genre	Rating
2 Days in the Valley	Comedy	R
2 Fast 2 Furious	Action	PG-13
2 Guns	Action	R
3 Days to Kill	Action	PG-13
3 Idiots	Comedy	PG-13
3 Ninjas	Action	PG

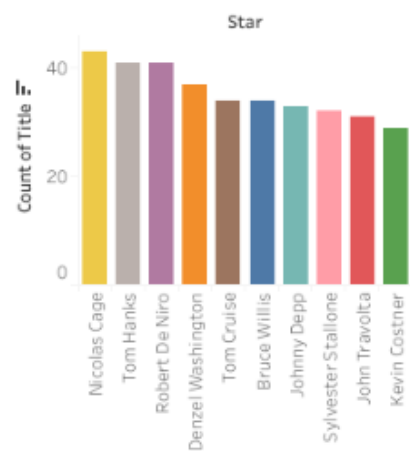
The Genre With The Most Movies




Company With Most Budget



Movie Star With The Most Movies



- 
- ▶ The Dashboard can be accessed via:
[Public Tableau](#)
 - ▶ Or can be downloaded via:
[GitHub](#)
 - ▶ Thankyou~

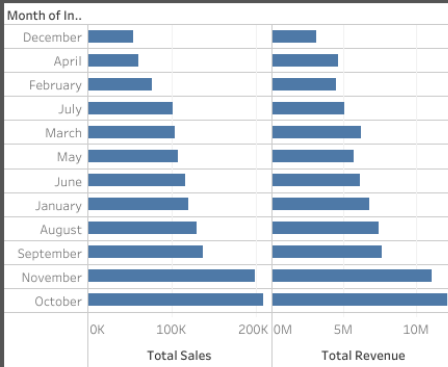
E-Commerce Dataset

- ▶ All the data utilized in this analysis are sourced from my Studi Independent Project. The Dataset itself can be access via [E-Commerce Dataset](#)

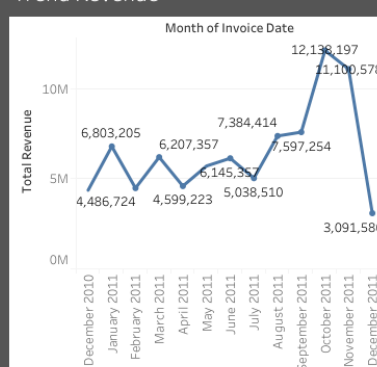
	A	B	C	D	E	F	G	H	I	J	K	L
1	InvoiceNo	StockCode	Product	Category	InvoiceDate	Quantity	UnitPrice(€)	Disc_type	CustomerID	Country	Market	Region
2	536370	POST	Reffair AX3C	Car & Motorbike	2010-12-01 08:45:00	15	18	1	12583	United States	US	East
3	536370	22726	Wayona Nyl	Computers & Accessories	2010-12-01 08:45:00	118	3.75	1	12583	Australia	APAC	Oceania
4	536370	21724	Ambrane Ur	Computers & Accessories	2010-12-01 08:45:00	106	0.85	3	12583	Australia	APAC	Oceania
5	536370	21913	Sounce Fast	Computers & Accessories	2010-12-01 08:45:00	96	3.75	1	12583	Germany	EU	Central
6	536370	21035	boAt Deuce	Computers & Accessories	2010-12-01 08:45:00	61	2.95	1	12583	Senegal	Africa	Africa
7	536370	22661	Portronics K	Computers & Accessories	2010-12-01 08:45:00	56	0.85	1	12583	Australia	APAC	Oceania
8	536370	22728	pTron Solerc	Computers & Accessories	2010-12-01 08:45:00	10	3.75	2	12583	New Zealand	APAC	Oceania
9	536370	22727	boAt Micro	Computers & Accessories	2010-12-01 08:45:00	106	3.75	3	12583	New Zealand	APAC	Oceania
10	536370	21883	MI Usb Type	Computers & Accessories	2010-12-01 08:45:00	96	0.65	1	12583	United States	US	West
11	536370	21791	TP-Link USB	Computers & Accessories	2010-12-01 08:45:00	10	1.25	2	12583	United States	US	South
12	536370	22326	Ambrane Ur	Computers & Accessories	2010-12-01 08:45:00	75	2.95	2	12583	United States	US	South
13	536370	22629	Portronics K	Computers & Accessories	2010-12-01 08:45:00	108	1.95	1	12583	Afghanistan	APAC	Central Asia
14	536370	22659	boAt Rugger	Computers & Accessories	2010-12-01 08:45:00	49	1.95	2	12583	Saudi Arabia	EMEA	EMEA
15	536370	22631	Portronics K	Computers & Accessories	2010-12-01 08:45:00	39	1.95	3	12583	Brazil	LATAM	South
16	536370	21731	Portronics K	Computers & Accessories	2010-12-01 08:45:00	122	1.65	1	12583	China	APAC	North Asia
17	536370	22900	MI Braided I	Computers & Accessories	2010-12-01 08:45:00	10	2.95	1	12583	France	EU	Central
18	536370	22540	Ambrane Ur	Computers & Accessories	2010-12-01 08:45:00	99	0.42	2	12583	United States	US	South
19	536370	22544	boAt Type C	Computers & Accessories	2010-12-01 08:45:00	56	0.42	2	12583	Italy	EU	South
20	536370	22492	Duracell USI	Computers & Accessories	2010-12-01 08:45:00	23	0.65	3	12583	Australia	APAC	Oceania
21	536370	10002	Flix Micro U	Computers & Accessories	2010-12-01 08:45:00	108	0.85	3	12583	Tanzania	Africa	Africa

E-Commerce Dashboard

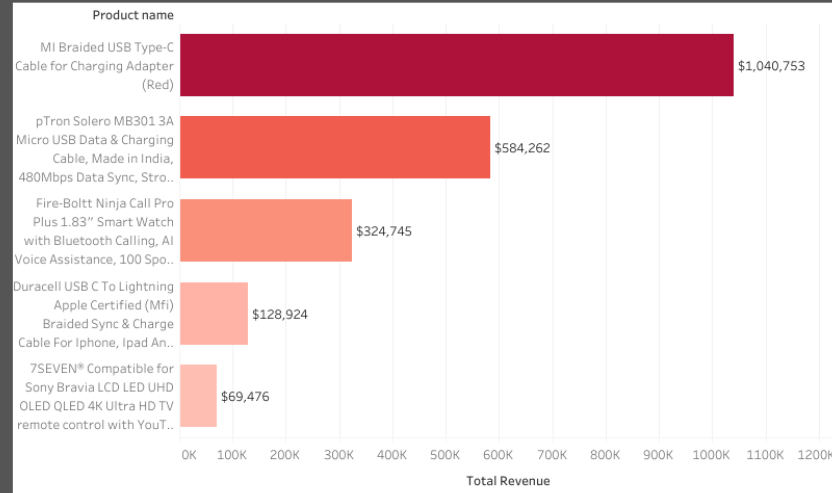
Total Revenue dan Total Sales



Trend Revenue



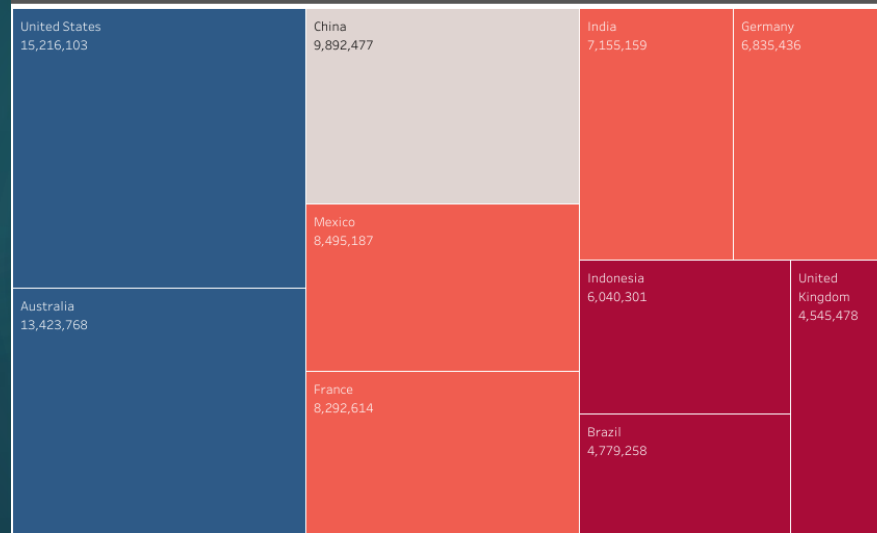
5 produk yang menghasilkan revenue paling tinggi pada rentang akhir tahun 2010 - 2011



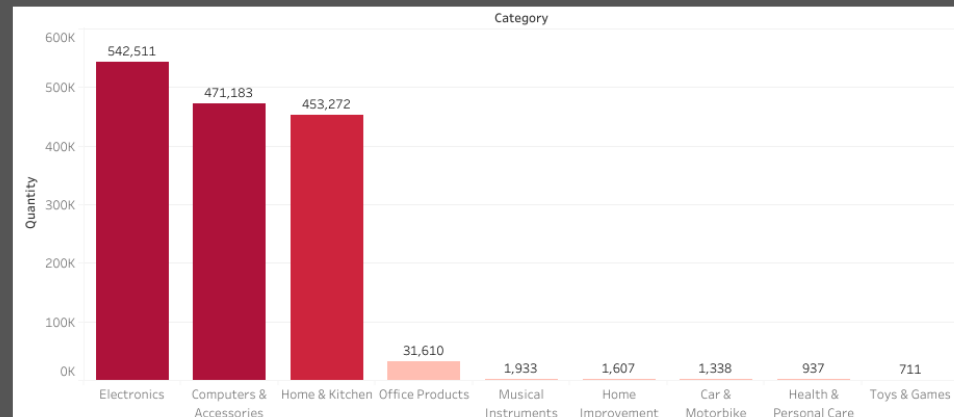
Month of Invoice Date
December 2010 to Decemb..


Region
All

10 Negara dengan Revenue Paling Tinggi

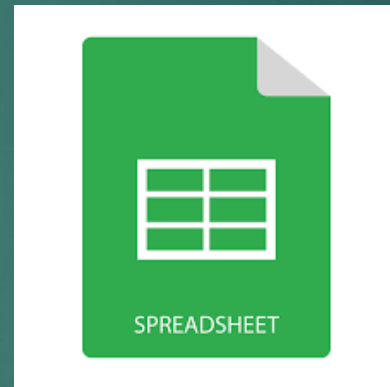


Kategori Produk Terlaris pada Tahun 2010-2011



- 
- ▶ The Dashboard can be accessed via:
[Tableau Public](#)
 - ▶ Or can be downloaded via:
[GitHub](#)
 - ▶ Thankyou~

Data Cleaning and Analysis



BY: RYNO PAHLEVI AL GHIFARI
RYNOPAHLEVIA@GMAIL.COM

Data Cleaning and Analysis -Kuala Lumpur Property Listings Dataset

- ▶ Performing Data Cleaning and Analysis on the Raw Kuala Lumpur Dataset, including:
 - **Duplicate Check:** Identifying and removing duplicate data.
 - **Data Cleaning:** Ensuring data quality and consistency through thorough cleaning processes.
 - **Data Types:** Verifying that each column has the correct and appropriate data type.
 - **Missing Value Check:** Identifying and handling missing values in the dataset.
 - **Outliers:** Detecting and managing outliers in the data.
- ▶ Subsequently, conducting the following analyses:
 - Calculating the average property price and categorizing the data into the first quartile (Q1), second quartile (Q2), and third quartile (Q3). Determining the mean, minimum, and maximum prices.
 - Identifying the categories with the most and fewest properties.
 - Determining the majority status of property furnishing.
 - Calculating the average property size.
 - Finding the location with the highest average property price.
 - Estimating the cost required to build a property with 5 bedrooms, 3 bathrooms, 4 parking spaces, and a lot size of 5000 sq ft.

- The Spreadsheet can be accessed via:
[Spreadsheet](#)

A1	Location								
	A	B	C	D	E	F	G	H	
1	Location	Price	Rooms	Bathrooms	Car Parks	Property Type	Size	Furnishing	
2	Bangsar, Kuala Lumpur	1,000,000.00	2	2	1	Condominium (Intermediate)	Built-up : 1,252 sq. ft.	Partly Furnished	
3	Bukit Bintang, Kuala Lumpur	1,000,000.00		2	1	Serviced Residence	Built-up : 1,012 sq. ft.	Fully Furnished	
4	Bukit Bintang, Kuala Lumpur	1,000,000.00	2	2	1	Serviced Residence	Built-up : 1,012 sq. ft.	Fully Furnished	
5	Bukit Bintang, Kuala Lumpur	1,000,000.00	2	2		Serviced Residence (Corner)	Land area : 1070 sq. ft.	Fully Furnished	
6	Bukit Jalil, Kuala Lumpur	1,000,000.00	3+1	3	2	Condominium	Built-up : 1,450 sq. ft.	Fully Furnished	
7	Bukit Jalil, Kuala Lumpur	1,000,000.00	3+1	3	2	Condominium	Built-up : 1,508 sq. ft.	Fully Furnished	
8	Cheras, Kuala Lumpur	1,000,000.00	4	4	2	2-sty Terrace/Link House (Intermed	Land area : 22x70 sq. ft.	Partly Furnished	
9	Cheras, Kuala Lumpur	1,000,000.00	4+1	5		2-sty Terrace/Link House	Land area : 22x80 sq. ft.	Unfurnished	
10	Cheras, Kuala Lumpur	1,000,000.00	4+1	4		2-sty Terrace/Link House (Intermed	Land area : 24x80 sq. ft.	Unfurnished	
11	Cheras, Kuala Lumpur	1,000,000.00	2	2	1	Serviced Residence	Built-up : 936 sq. ft.		
12	City Centre, Kuala Lumpur	1,000,000.00	5	6	1	Condominium	Built-up : 1,830 sq. ft.	Partly Furnished	
13	Damansara Heights, Kuala Lumpur	1,000,000.00	1	2	1	Serviced Residence	Built-up : 999 sq. ft.	Fully Furnished	
14	Dutamas, Kuala Lumpur	1,000,000.00	2	2		Serviced Residence	Built-up : 1,145 sq. ft.	Fully Furnished	
15	Dutamas, Kuala Lumpur	1,000,000.00	2	2	1	Serviced Residence	Built-up : 1,145 sq. ft.	Fully Furnished	
16	KL City, Kuala Lumpur	1,000,000.00	3	3	2	Serviced Residence	Built-up : 716 sq. ft.	Partly Furnished	
17	KLCC, Kuala Lumpur	1,000,000.00	3+1	3	2	Serviced Residence (Corner)	Built-up : 1,313 sq. ft.	Partly Furnished	
18	KLCC, Kuala Lumpur	1,000,000.00	1	1	1	Condominium (EndLot)	Built-up : 657 sq. ft.	Fully Furnished	
19	KLCC, Kuala Lumpur	1,000,000.00	Studio	1		Condominium (Studio)	Built-up : 657 sq. ft.	Fully Furnished	

Exploratory Data Analysis Using Python



BY: RYNO PAHLEVI AL GHIFFARI

RYNOPAHLEVIA@GMAIL.COM

EDA NYC Property Dataset Using Python

- ▶ Dataset itself is from My Studi Independen that can be access via [NYC Property Dataset](#)
- ▶ Performing data preparation, data cleaning (such as data imputation, data cleaning, and outlier checks), and conducting Exploratory Data Analysis (EDA) with the following objectives:
 1. **Average Price, Cheapest, and Most Expensive:** Analyzing the average price, cheapest, and most expensive properties for each room type, along with their characteristics.
 2. **Total Properties and Average Price per Neighborhood Group:** Calculating the total number of properties and the average price per neighborhood group.
 3. **Average Reviews and Availability (2017-2019):** Determining the average number of reviews per month and the average availability of rooms for each room type per year from 2017 to 2019.

- ▶ The Google Colaboratory can be accessed via:
[NYC Property - Google Colab](#)

The screenshot displays the Google Colaboratory interface. On the left, a sidebar contains a 'Table of contents' and a 'Prep' section with icons for data imputation and cleaning. The main workspace is divided into two sections: 'Code' and 'Text'. The 'Code' section contains two code cells. The first cell imports necessary packages: pandas, numpy, and matplotlib. The second cell mounts the Google Drive. Below the code cells, a message indicates that the drive is already mounted. The 'Text' section contains a code cell that reads a CSV file from the mounted drive. The output of the code cells is visible at the bottom, showing the shape of the data and the first 10 rows.

Table of contents

Prep

Imputasi Data

- Imputasi data kosong pake kolom 'name'
- Imputasi data kosong pake kolom 'host_name'
- Imputasi data kosong pake kolom 'price'
- Imputasi data kosong pake kolom 'last_review', 'number_of_reviews', 'reviews_per_month'

Cleaning Data

- Cek duplikat
- Cek kolom

neighbourhood_group

minimun_nights

Outliers

Insert code cell below
Ctrl+M B

```
[ ] #Import Package yang diperlukan
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

```
[ ] #mount dengan google drive
from google.colab import drive
drive.mount('/content/drive')
```

Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True).

```
data = pd.read_csv ('/content/drive/MyDrive/Studi_Independent/Modul 4/AB_NYC_2019.xlsx - AB_NYC_2019.csv')
```

Data Preparation

```
[ ] data.shape
```

```
(5018, 15)
```

```
[ ] data.head(10)
```


Clustering Laptop Customer Using Python

- ▶ The Dataset itself is from My Studi Independen that can be access via: [Customer Laptop Dataset](#)
- ▶ Performing data preparation and data cleaning, including data imputation, data cleaning, and outlier checks. Then, conducting clustering analysis with the following methods:
 - **K-Means Clustering:** Applying the K-Means algorithm to identify distinct clusters within the data.
 - **Cluster Cardinality:** Determining the number of elements in each cluster.
 - **Cluster Magnitude:** Analyzing the magnitude and characteristics of each cluster.

- The Google Colaboratory can be accessed via:
[Customer Laptop Google Colab](#)

The screenshot displays the Google Colaboratory web interface. On the left, a sidebar contains a 'Table of contents' and a search bar. Below these are sections for 'Prep' (with a search icon), 'Imputasi Data' (with a list icon), and 'Cleaning Data' (with a list icon). The 'Imputasi Data' section lists four tasks: 'Imputasi data kosong pake kolom 'name'', 'Imputasi data kosong pake kolom 'host_name'', 'Imputasi data kosong pake kolom 'price'', and 'Imputasi data kosong pake kolom 'last_review', 'number_of_reviews', 'reviews_per_month''. The 'Cleaning Data' section lists two tasks: 'Cek duplikat' and 'Cek kolom'. Below these are two more sections: 'neighbourhood_group' and 'minimun_nights'. The 'Outliers' section is partially visible at the bottom.

The main area of the interface shows a Jupyter Notebook with the following code cells:

```
[ ] #Import Package yang diperlukan
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

```
[ ] #mount dengan google drive
from google.colab import drive
drive.mount('/content/drive')
```

Below the code cells, a message states: "Drive already mounted at /content/drive; to attempt to forcibly remount, call drive.mount("/content/drive", force_remount=True)." This is followed by a code cell that reads a CSV file from Google Drive:

```
data = pd.read_csv ('/content/drive/MyDrive/Studi_Independent/Modul 4/AB_NYC_2019.xlsx - AB_NYC_2019.csv')
```

The notebook also includes a 'Data Preparation' section with the following code cells:

```
[ ] data.shape
```

```
(5018, 15)
```

```
[ ] data.head(10)
```