**DASHBOARD**

**Product Name: Azure Cleaning Engine**

**Product Version: Version 1.0.0**

**Product Phase: Sprint 4**

**Current Date: 2022-07-02**

[**Context 👀** 1](#_Toc103077368)

[**Problem Statement** 1](#_Toc103077369)

[**Solution Statement** 2](#_Toc103077370)

[**My Approach** 2](#_Toc103077371)

[**Research Question** 2](#_Toc103077372)

[**Product Dashboard** 3](#_Toc103077373)

[**Conclusion and reflection on this assignment** 4](#_Toc103077374)

# **Context 👀**

This dashboard is designed to provide multiple insights into the New York dataset.

The dashboard helps the user or data analyst identify trends more easily and get the right insight into the Big Data cab in New York. It highlights the opportunity to understand customer behavior and improve business decisions for taxi company. Here I will use the Tableau dashboard prep, analyze, collaborate, form a interactive business vizualiation from New York Big Dataset.

# **Problem Statement**

After data processing, we have came up with clean raw dataset. Customers are not able to find and understand their data insight. With raw big data, they are missing a tool to visualize these actual data from cloud or local environment. Without the dashboard, it gains trouble for client or stake holder to find out their business data performance. They cannot instantaneously see spikes, download slopes, and other trend that indicate success or failed, tipped or non-tipped driver.

Hence, they cannot explore the historical data. It might lead to another problem; the company taxi cannot be aligned and hard to achieve their goal in business. Such know where to increase the number of taxi or increase the working hour of taxi driver.

# **Solution Statement**

After the research, I have found that there are not so many tool to create the interactive dashboard and do not provide all the power of a data counter. This means that it will be hard to see the trend of New York City's monthly historical data to give the right meaning to the data.

My solution is the actual dashboard that allows you to visually display the data. It clarifies the need to understand datasets through a graph and a visual diagram of the customer behavior. All the charts in the dashboard explain the different meanings of large data sets. Here I will use Tableau application to create an interactive dashboard. It will contain more than 6 graphs including GIS (Geographic Information System) mapping the dataset to view the performance of cabs in New York from a helicopter perspective.

# **My approach / Method**

First, I approach to analyze the big data source from NYC Datahub. After getting through a whole process of cleaning and preparation data from previous assignment in Python. I export those cleaned dataset to store in a local file.

Aforementioned, taxi company needs to find the data insight. I will use cleaned dataset to create the dashboard being more interactive, I also add an external dataset to include map coordination to show the customer location and movement taxi from point A to point B. It merges both dataset from Weather Data, and Zone of Taxi Data. There will multiple boroughs present for different district in New York Megacity. We can state them into 6 district include: Manhattan, Brooklyn, Bronx,,..ect. Then, I am able to be visualized them in history data and Geographic Visualization to analyze the taxi and customer behavior via map technology.

# **Research Question**

*What kind of useability that I used for the visualization of data?*

* Dashboard analytic help Uber’s Business to give more detail of data insight. Within the current tableau dashboard, I can help Uber find out which is current of customer trend and taxi driver behavior. Therefor, Uber can use the information which has been analyzed and boost their business decision.

*How large datasets of structured data be precisely visualized to identify the correct data trend ? If so, what kind of trend can be identified with your solution?*

Yes, It possible to manipulate a large dataset and analyze them with a dashboard. Dashboard should be large and enable scaling and expand the complexity of data graph. With large resolution dashboard, we can add the GIS dataset and visualized them in different era of the city, it will let the customer know the active zone of taxi driver and user activity. Dashboard can also be able to extend with cloud data source so we can spot out more various necessary trend with high accurately level depend on our feature choice. We can find some important trend of data impact and affect to customer behavior such as:

* Number of passengers per weekday.
* Trend of Taxi Analysis.
* Top pickup district where most customer took the taxi driver.
* Trending of Tipped Driver.
* Map of Density Active Zone.

All trend which mentioned above allow Uber – Driver Company control and know where to boost their taxi trip and improve their business performance.

# **Product Dashboard**

Graphical user interface, application, map

Description automatically generated**Dashboard Tableau: show me the detail of business Uber Taxi in New York.**

**Figure 2 Tableau Dashboard**

This dashboard represents the behavior of customers and their movement from pickup point to pickup point. From left to right they have a logical way to show the number of trips distance. And analyze the trend of the number of cab trips from Monday to Friday. So the data analysts know where and when the habits of customers change to use a cab on a weekday. In the 3rd graph in the right corner, the company learns how many passengers travel per weekday.

The 2 GIS map allows dashboard to explore the ability of visualization the density of cab trips in the city, and it is possible to show the busiest zone in the borough of Manhattan. They use the Google API from the back-end of this dashboard. I connected to the Geo Location Map API to draw this map in two ways. One is the satellite view and the other from the bottom corner of this image is a traffic-based map. At the top of the map, I applied the district filter to show where we can filter districts from various related graphics.

# **Conclusion and reflection on this assignment**

Top impact of this dashboard useability to business:

1. Its understandable for data analyst to find out data insight.
2. It’s interactive with map of taxi.
3. It’s sharable. Therefore, other business analyst possible to see.
4. It lead to the business approach and respond quickly update the trend of data.

Developing an interactive dashboard takes time and patience. The most difficult part I learned is how to use Cloud's API to visualize the mapping part. The second hurdle is creating an appropriate filter for different criteria, such as a filter for the district that shows a different number of driving miles or tipping percentages. Giving the right title and filter function and telling the logical data story from the dashboard is a journey of searching and understanding the structure dataset to create the right graph that the customer needs. With Big Data, I need a good background in data processing, which can be found in the previous task from the cleaning notebook. The resource output is in the form of JSON or CSV files. The resources play an important role in connecting and creating a good dashboard.

I realized that the lesson to creating a good dashboard is to create a wireframe before creating a dashboard. I learned that it is more important to create a proper chart is more clearly compare with a random visualization and developing time is short to create the dashboard. To avoid wasting time when creating dashboards, I tested some dashboard technique before apply Tableau.

My conclusion is that this dashboard is capable of adding more data sources from NYC and updating or resetting in a timely manner. Our dashboard can be further expanded and enlarged more graph following the need of the client or business analyst. The Uber company will use this dashboard for forecasting and better business decisions. The product improves the company's knowledge of its customers' rides and helps the business customer make decisions to improve its service.

**Reference**

*New York Data Set – Retrieve from: https://opendata.cityofnewyork.us/*

*Databrick – Retrieve from:* [*https://docs.microsoft.com/en-us/azure/databricks/scenarios/what-is-azure-databricks*](https://docs.microsoft.com/en-us/azure/databricks/scenarios/what-is-azure-databricks)

*Video of Product is included of the assignment. Linked from the submission.*