



## User Guide for Analyst

# Envisioning Yellow Taxi Demand in NYC City

---



**Submitted by: Simranjeet Randhawa**

**Student ID: 200412297**

## Table of Contents

1. Synopsis .....	1
2. Accessing Site.....	1
2.1 Link to Site.....	1
3. Accessing the Cloud Portal .....	1
3.1 Credentials are important for Analyst.....	1
3.2 Google Cloud Architecture for Analyst .....	3
4. How to create more dashboards .....	3
4.1 Sharing with other Analyst .....	4
5 How To Understand And Interpret Performance Of Model .....	5
5.1 Feature Importance .....	5
5.2 Model Accuracy Measures.....	5
5.3 How to Use the Model .....	6
6 Creating New Measures:.....	7
6.1 Previous : .....	7
6.2 How to add Create New Measures:.....	7
6.3 Analyst can add visuals:.....	8
6.4 Analyst have multiple options for visualization: .....	8
6.5 Give a Final Touch up to the Dashboard: .....	9
7. Summary .....	10

## 1. Synopsis

The Envisioning Yellow Taxi demand in NYC is a tool that is very useful for the yellow cab taxi driver to find the number of pickups in a particular area. This document will serve as a guide on how to use this tool and focused on analyst.

## 2. Accessing Site

Use the URL given below to access the site:

### 2.1 Link to Site

<https://deploy-flask-266818.appspot.com/>

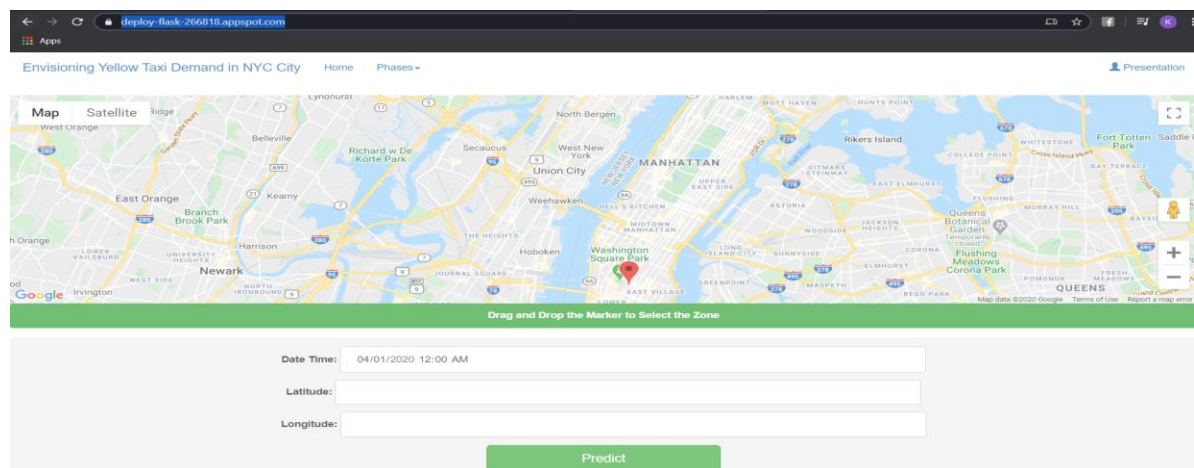


Figure 1: Displays home page

## 3. Accessing the Cloud Portal

### 3.1 Credentials are important for Analyst

- The analyst can get the Google Cloud credentials from the [Service Account Page](#).
- Once the analyst acquires the credentials. The credentials can be used to monitor the performance of the machines on the Google Cloud.

- The below gives the Dashboard of my project that monitors all the devices and API that are made to my deployed model. An analyst can use the credentials to get insights into the resources used. Also, it gives the cost used.

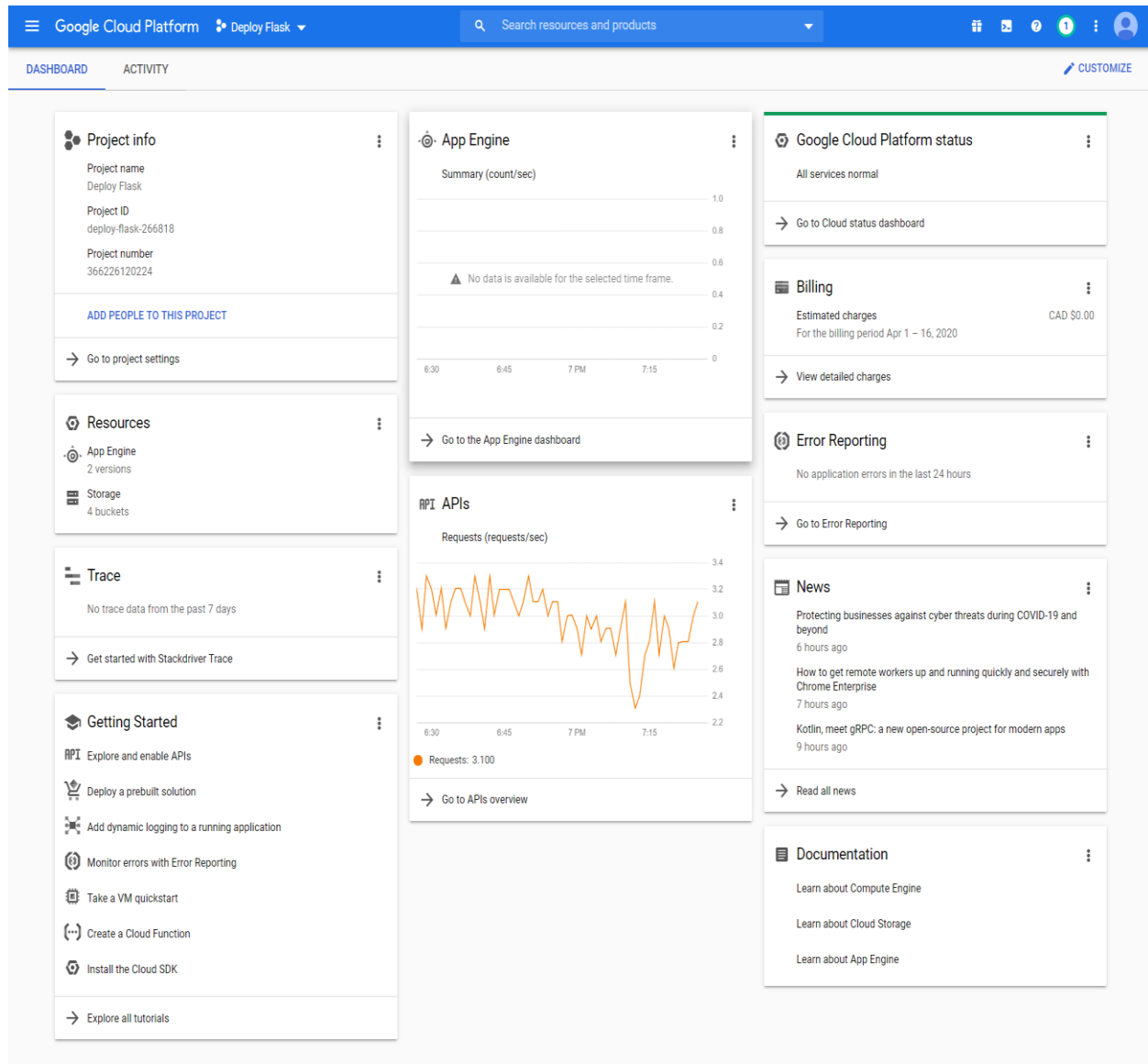


Figure 2: Dashboard that can track the details

Also, the Google Cloud Platform gives the ability to customize the Dashboard to add more visuals that can forecast the price of the deployed application. Moreover, another interesting thing the analyst can use Google Cloud platform to keep track of activity.

Google Cloud Platform <span>Deploy Flask</span> <span>Search resource</span>			
DASHBOARD	ACTIVITY	<span>X</span> HIDE FILTER	
3:53 AM	Completed: Create VM	366226120224@cloudservices.gservi...	▼
3:52 AM	Create VM	366226120224@cloudservices.gservi...	▼
		Filters	CLEAR

Figure 3: Activities Logs can be checked by analyst

### 3.2 Google Cloud Architecture for Analyst

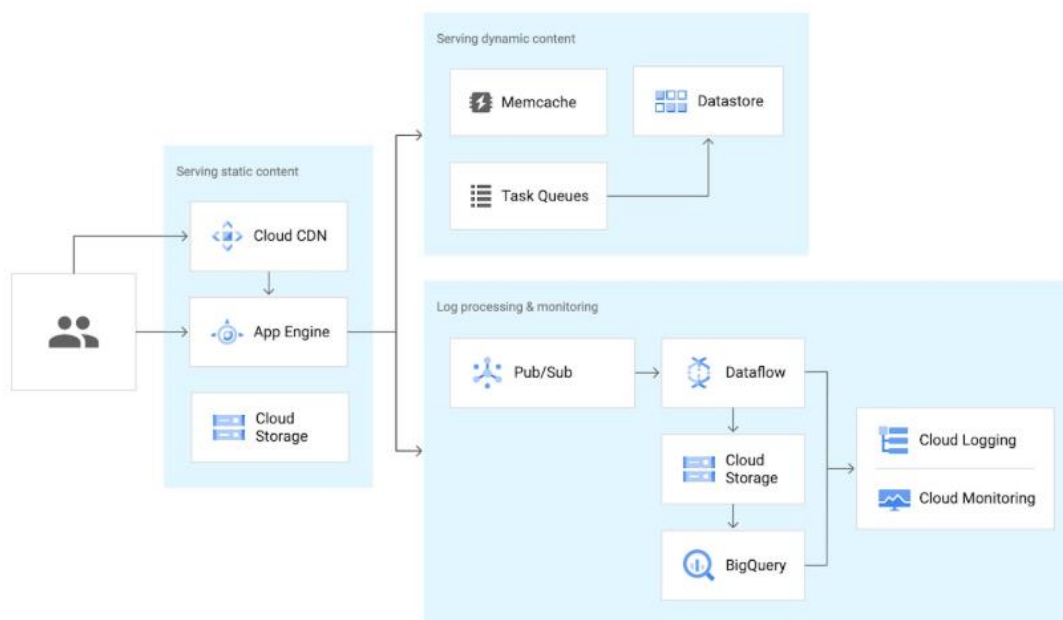


Figure 4: Data Flow Diagram using App Engine

The above figure depicts the flow diagram that the App Engine adopt which will be useful for the analyst to study the relationship.

## 4. HOW TO CREATE MORE DASHBOARDS

The following steps can be used to create more dashboards:

- Go to <https://datastudio.google.com/navigation/reporting> and select a blank report.

- Click on Add Data to Report. Select Big Query it will give access to the project tables. But please note it is essential to set up the data on the cloud initially following step 1.

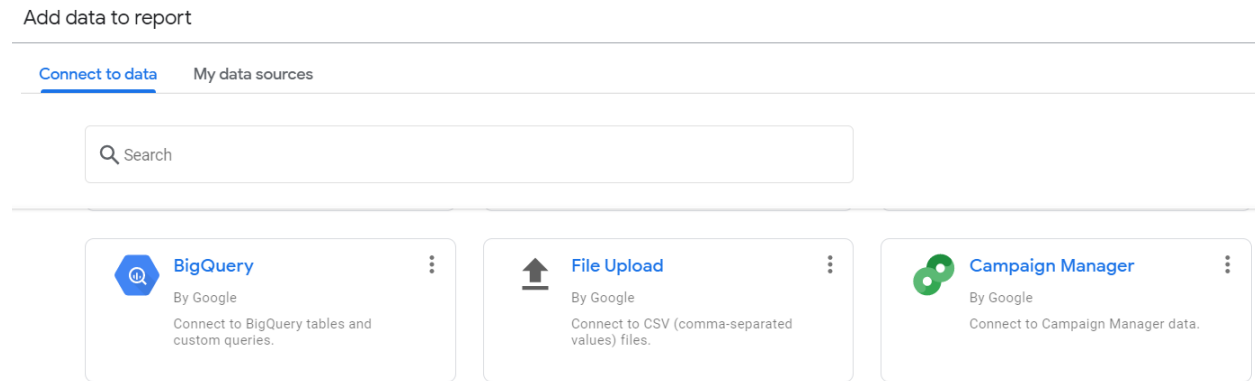


Figure 5: Connection to Data

Apart from this an easy method can be clone my dashboard and start scaling my dashboard. I can easily share and permit to edit the dashboard which can then be edited by any of the analysts. To request for permission please send me an email: [ssrbazpur@gmail.com](mailto:ssrbazpur@gmail.com)

### 4.1 Sharing with other Analyst

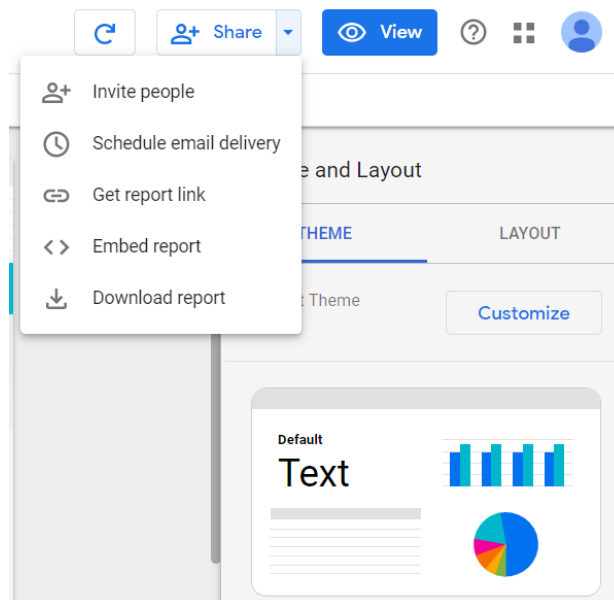


Figure 6: Sharing Report with other team members

If new dashboards are to be made instead of following steps 1 and 2 the analyst can clone my previous dashboard and add a page in the report to create explore new insights in the updated table. I can also invite other team members of the analyst and everyone can work in parallel with the use of Data Studio.

## 5 HOW TO UNDERSTAND AND INTERPRET PERFORMANCE OF MODEL

Also, I have extracted the important feature that the model considers:

### 5.1 Feature Importance

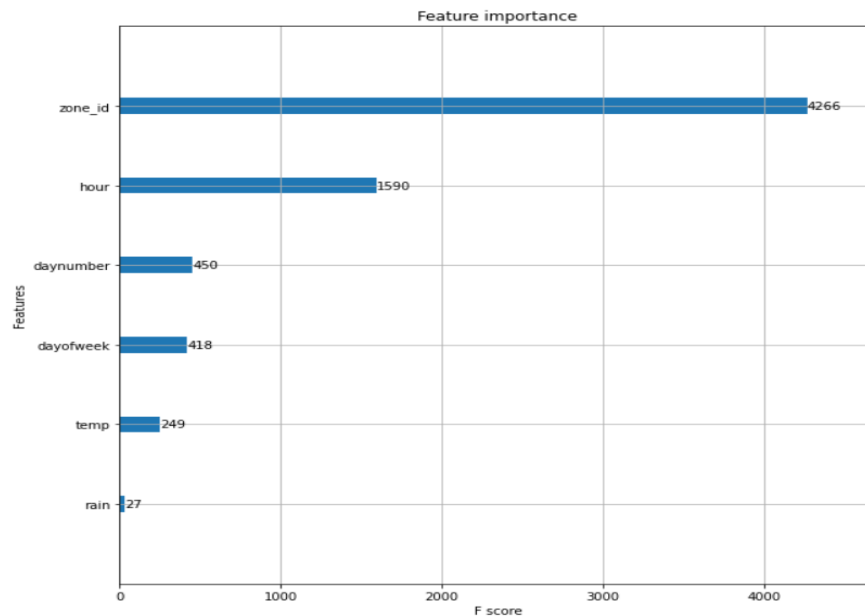


Figure 7: Important Features

### 5.2 Model Accuracy Measures

The model used was Random Forest Regressor. As it is a regression-based model so the performance of the model was measured using the following measures:

Model	MAE	R2	RMSE
Random Forest Regressor	11.57897810134032	0.45528366614765486	24.93736930641297
Xgboost	13.57897810134032	0.7528366614765484	32.93736930641297
Light GBM	18.57897810134032	0.9552836661476548	40.93736930641297
ARIMA	14.95411850134032	0.81357330134032	35.56811210134032
Linear Regression	21.57677810134032	0.9972836661476548	43.93737930641297

Figure 8: Model Error Measures

The below gives a bar plot showcasing the different error measures for the different models tested. An analyst can easily get quick insights while using the models below.

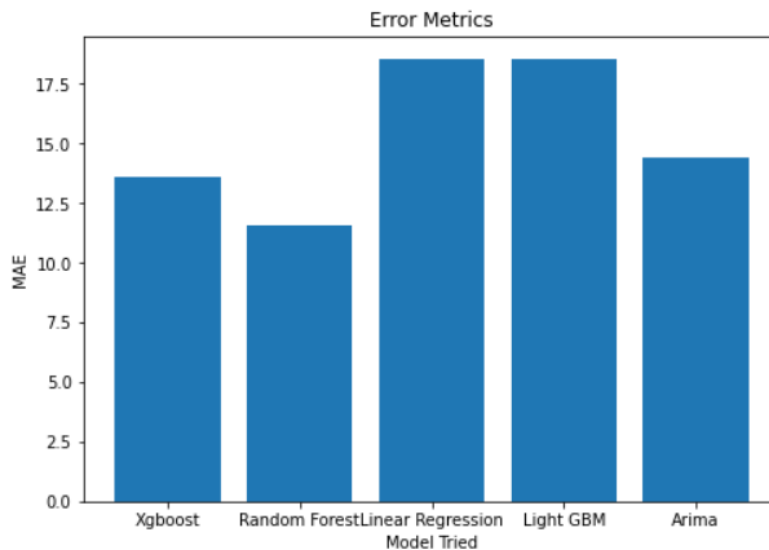


Figure 9: MAE for the models

I also calculated the MASE for the models that give the least MAE and then got to know the best model possible.

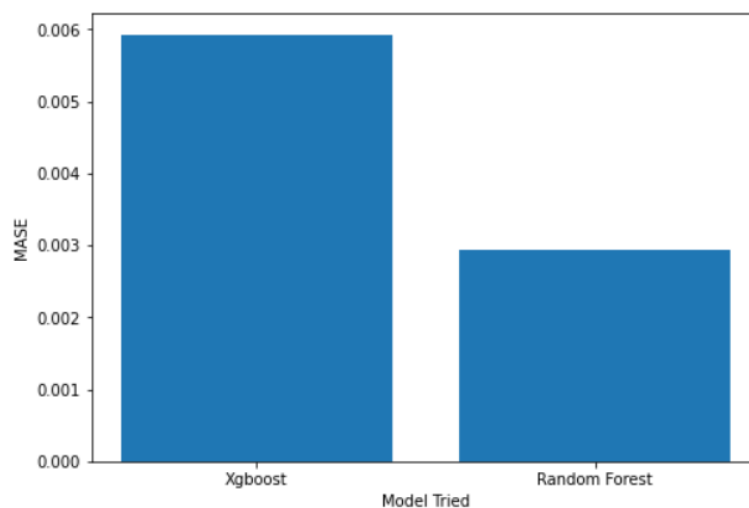


Figure 20: MASE of models

The above bar chart analysis can help the analyst know the model evaluation parameters.

### 5.3 How to Use the Model

To use the model the pickle file can be used by the analyst to build other applications.



6 Creating New Measures:

6.1 Previous :

Yellow Taxi Demand			Data From Google Cloud Bucket
Average Pickups per Hour	Average pickups per Day	Average Pickups per Week	Week Total Trips
12,838	308,123	2,156,860	2.2M

Figure 31:

These are the metrics that are currently calculated in the Forecasted Dashboards. An analyst can easily add new measures.

6.2 How to add Create New Measures:

There is the following on the right-hand side of the:

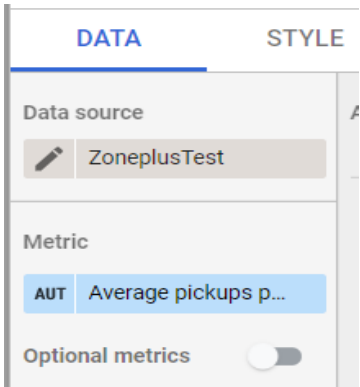


Figure 42:

On the Metrics click on to the metrics field to create metrics than enter the Data Studio formula to create a new measure. Like instance the below shows a metrics with the average pickups:

AUT Average pickups per Day

Name

Average pickups per Day

Formula

1

SUM( Total Trips )/COUNT\_DISTINCT( DayNumber )

Figure 53:

### 6.3 Analyst can add visuals:

Like for instance below is a heat map showing the top 10 demanding zones for the upcoming week.

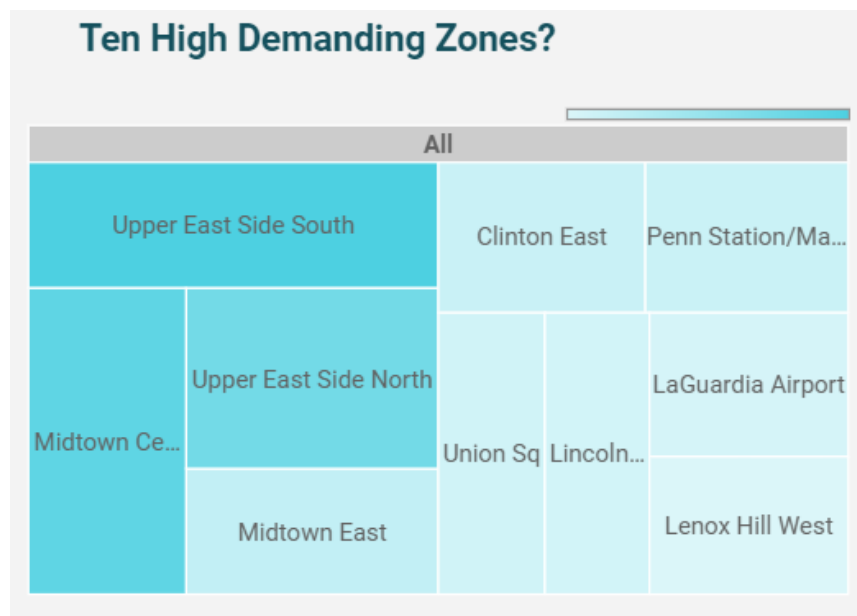


Figure 64:

### 6.4 Analyst have multiple options for visualization:

These are the sample charts that are available to the analyst in Data Studio.

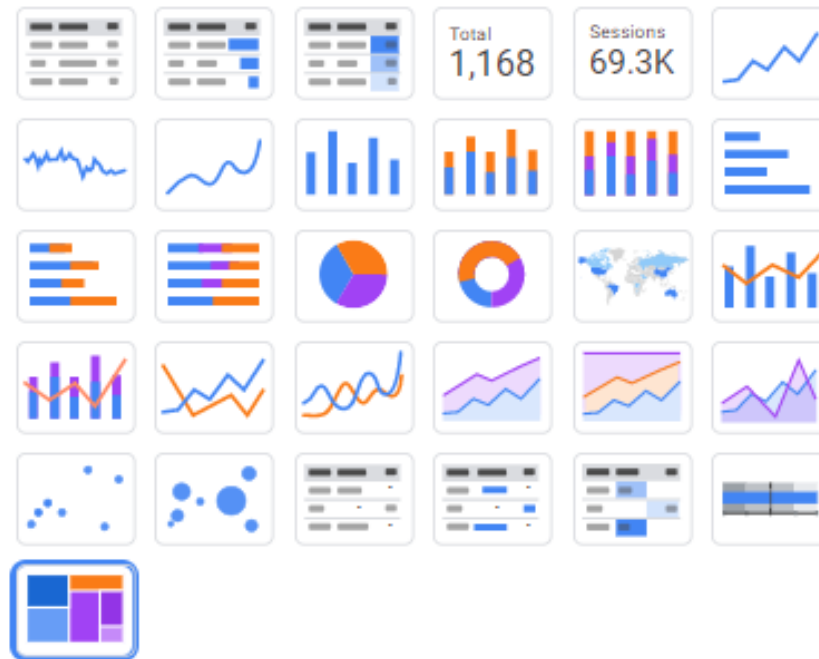


Figure 75:

### 6.5 Give a Final Touch up to the Dashboard:

Multiple inbuild themes provide the ability to create a dashboard



Figure 86:

## 7. Summary

Model Used – Random forest regressor.

Due to the least – MASE, R2, and RMSE.

Model Compared- Xgboost, ARIMA , LightGBM, Linear Regression.

-----X-----X-----