**Problem Statement**- The occupancy rate of the Airbnb is close to 32%. This presents us with the problem if we can optimize the revenue based on the occupancy rate or if there is way to increase the occupancy rate.

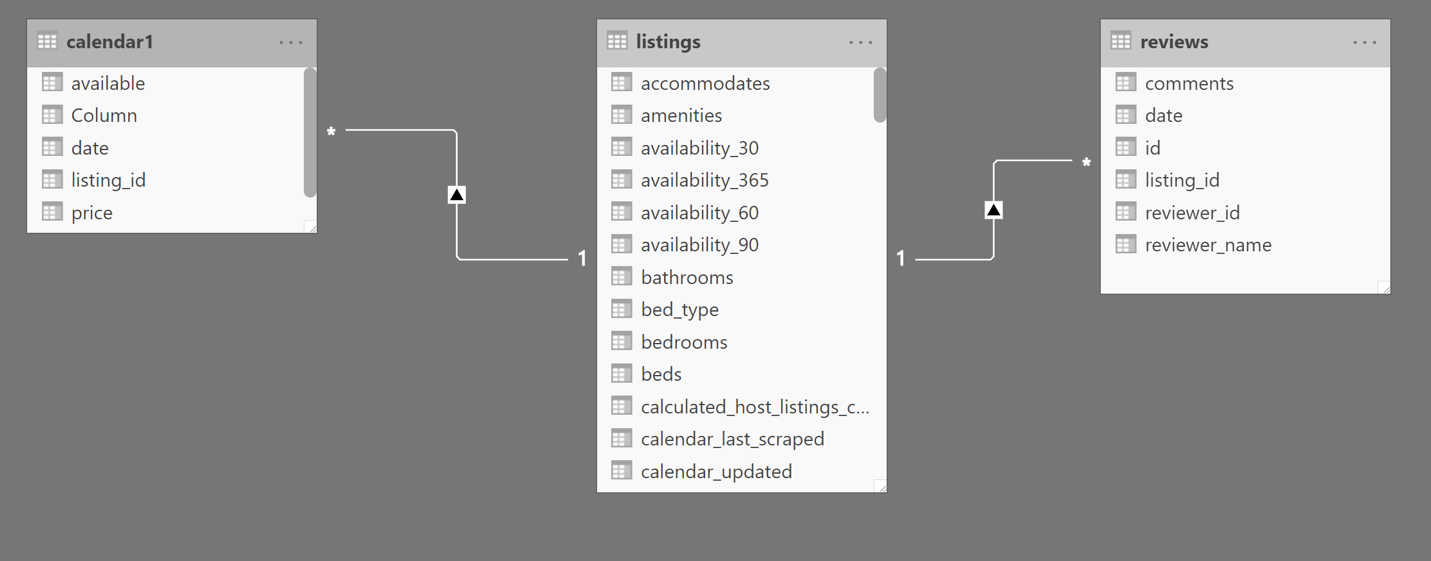
**DATASET- The dataset has been taken from Kaggle.** The dataset describes the listing activity of homestays in Seattle, WA. The dataset has three parts

1. Listings- Listings, including full descriptions and average review score. Reviews, including unique id for each reviewer and detailed comments
2. Calendar- Calendar, including listing id and the price and availability for that day
3. Reviews- Contains the reviews of the property

The following the descriptions of the columns in the dataset

|  |  |
| --- | --- |
| **Column Name** | **Description** |
| listing\_id | Unique Id for every listing on the website |
| name | Name of the property |
| summary | Brief Introduction to property |
| description | Compelete descrption of the property |
| transit | Nearby Transist available description |
| host\_id | Unique Id for every host on the website |
| host\_name | Host Name |
| host\_since | Years host has been listed on the website |
| host\_location | Host Location |
| host\_about | Brief Introduction to host |
| host\_response\_time | Reponse rate for the host |
| host\_acceptance\_rate | Acceptance rate for the host |
| host\_is\_superhost | To decribe if the host is super host or not? |
| city | City |
| state | State |
| zipcode | Zip Code for the property |
| country | Country |
| accommodates | Number of people allowed |
| bathrooms | Number of bathrooms at the property |
| bedrooms | Number of bedrooms at the property |
| beds | Number of beds per bedroom |
| bed\_type | Type of bed available at the property |
| amenities | Amenties include |
| price | Per day price |
| weekly\_price | Per week price |
| monthly\_price | Per Month price |
| security\_deposit | Security Deposit |
| cleaning\_fee | Cleaning fees |
| guests\_included | Guest included per booking |
| extra\_people | Number of extra guests that can be accomodated in the room |
| minimum\_nights | Minimum stay in days |
| maximum\_nights | maximum stay in days |

**Schema-**

****

**Use cases-**

1. Increase occupancy rate
2. Add new listing in the popular

When the user search for the accommodation at new place, the user gets the search results based on the most the criteria which highly matches the user search criteria. So, there might be loss of revenue if the user which is willing to pay more gets the results that matches almost their criteria, or the user doesn’t find the best fit that matches their budget.

We aim to solve the problem based on the flexibility of the budget that the user has.

We will provide two kinds of search on the basis of the traveler being on budget or If the user is willing to pay more.

We have segmented our customer based on the Flexibility of the budget

1. Excursionist- To not loose on the potential revenue from the rich client, we optimize the search result to give the user feature rich accommodation. This will increase revenue based
2. Budget Travelers- We don’t want to ward off the budgeted travelers by giving the search result that doesn’t fit their pocket. So we have the recommendation based system in place to provide search to provide maximum facilities with minimum price.

**W5H**

**Why-** To capture the untapped market and to increase the market penetration index for the company. The increase in market penetration would ultimately increase the revenue of the company. The dashboard will provide insights into the segmentation and our strategy to increase the revenue.

**What-** We are creating a hotel segmentation based on the budget criteria. We have the feature rich accommodation for the luxury travelers and budgeted options for the tourists on restricted budget options.

**When-** We have our interaction with customers through the website. We aim to entice our customers with tiered pricing and rule-based pricing. We try to project the company’s aim to have greater market penetration based on attracting more customers.

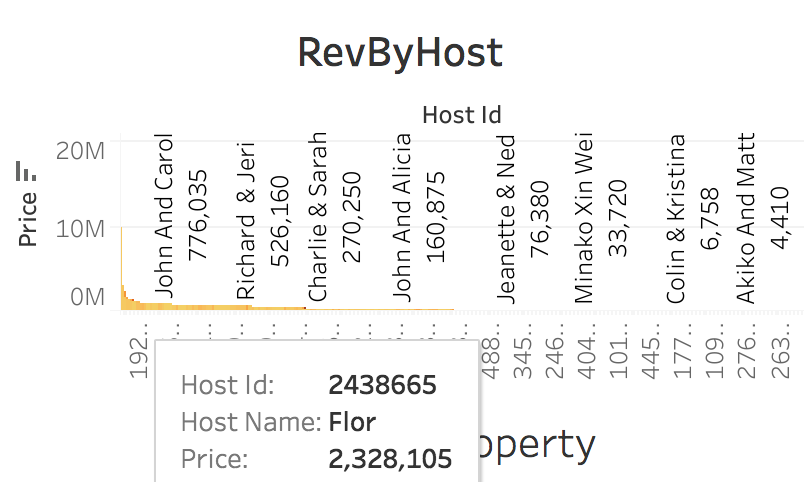
**Where-** The customer has the choice to choose if he is on restricted or if he wants the maximum features for their accommodation.

**How-** We will market the entire product based on two factors

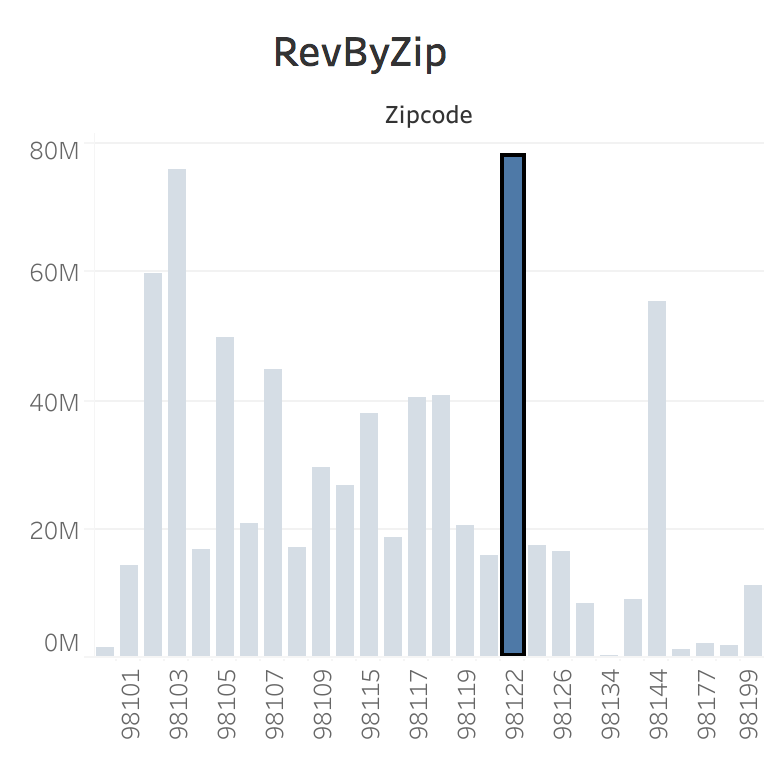
1. Pricing strategy
2. Hotel segmentation

**KPI’s-**

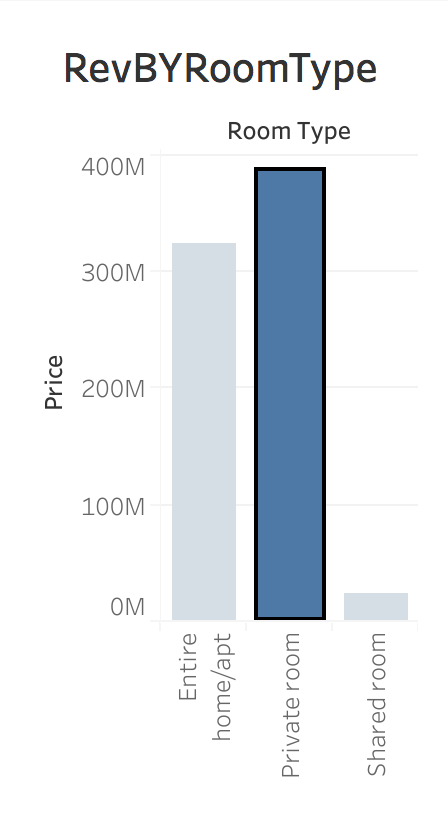
1. **Revenue By Host-** This will indicate how a particular host is performing. The host that has the most revenue is the one that is most popular and that can include in our search results. Currently Flor is the best performing host as per the revenue generated.



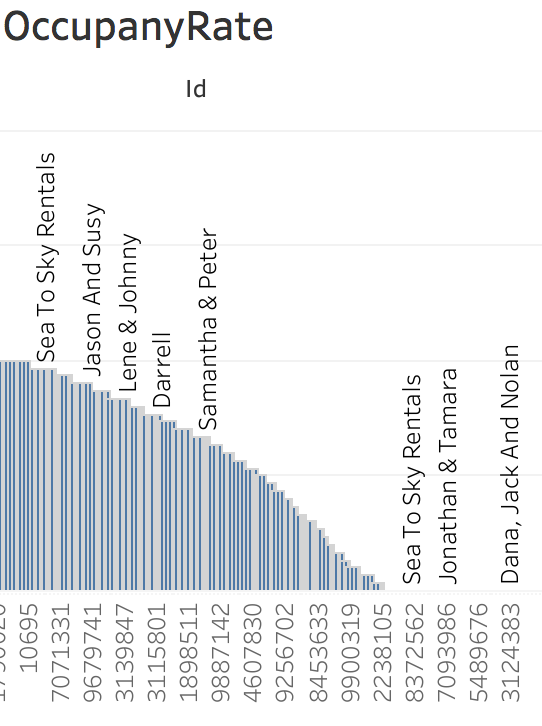
1. **Revenue by Zip Code-** This will indicate the most popular area. This is where more tourist would like to stay. Currently zip code 98112 is the best revenue generating zip code. We can also include more listings within the popular zip code.



1. **Revenue by Room Type-** It shows which type of room is currently more in demand and which will generate more profit operating cost. Currently private rooms are the most popular type of the room with $385273534 in revenue.



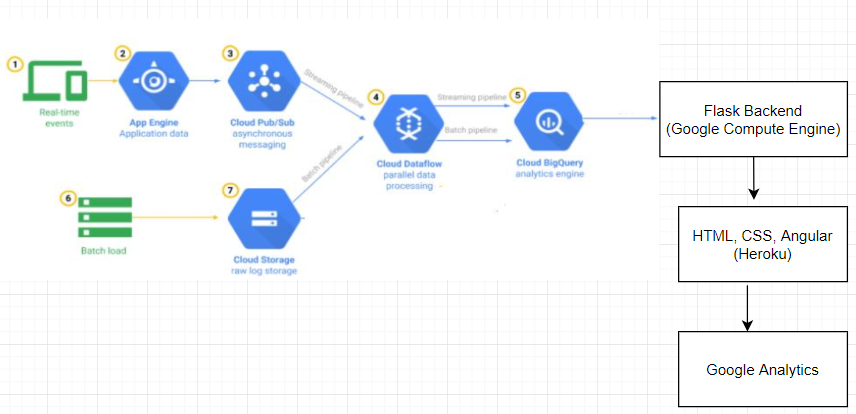
1. **Total available rooms** – We have in total of 3819 rooms available across Seattle.
2. **Average Daily Rate**- Average daily price paid per room across Seattle is $127.
3. **Revenue Per Available Room (RevPar)-** This accounts for the average daily rooms revenue generated per available room. Increasing RevPAR means either that rates or Occupancy Rate are rising, or both
4. **Average occupancy rate-** Occupancy is a percentage of the available rooms occupied for a specific period. The Occupancy rate for Airbnb over the year is 32% with January having the occupancy rate as 36% as the best performance and December having the share of 27% as the worst performance. Usually, the higher the occupancy the better because the company is earning more revenue than companies with low occupancy. However, this may not always hold true if the company cuts prices to boost its occupancy. The rate is also key to the operational side of the business to ensure proper staffing and inventory.



1. **Gross operating profit per available room(GOP PAR)-** This factors the operating revenue minus the operation cost. With this the Hoteliers are able to see profit across all revenue centers/the sum of all the parts- not just rooms. The average revenue generated is $140.45 with 699883 rooms available to generate the $98299455 over the year.
2. **Market Penetration Index (MPI) or Occupancy Penetration Index**- This hotel performance metric measures how a hotel’s occupancy compares to a competitive set. The metric serves as a guide to understanding a hotel’s dominance and demand in the marketplace. Increasing demand, naturally leads to more revenues. The MPI for AIRBNB stands at 32% over the year.

#### Average Rate Index (ARI)- This hotel performance metric measures how a hotel’s average daily rate compares to a competitive set. This metric serves as a metric to pricing right in the marketplace as well as an illustration of a hotel’s rate performance against its competition, helping determine whether rates needs to be lowered or elevated. The Average Rate stands at 0.94 which indicates that the room rates are priced right.

# Architecture-

****

**Pipeline and deployment:**

Programming Language used: Python

Operating System: Windows, Linux

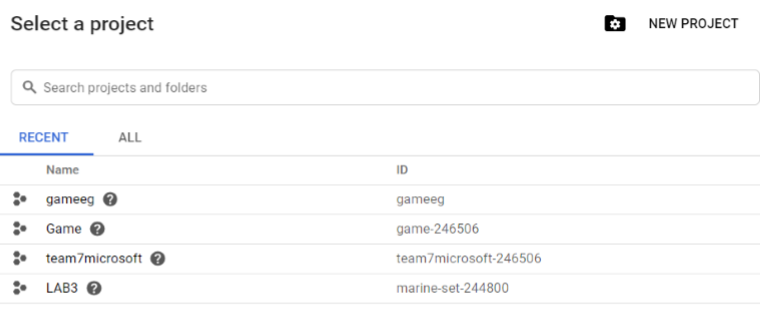
**Tasks**:

1. Getting the required data from the Google Cloud Platform in real time
2. Exploratory Data Analysis
3. Data pre-processing and cleaning
4. Developing solutions for business problem

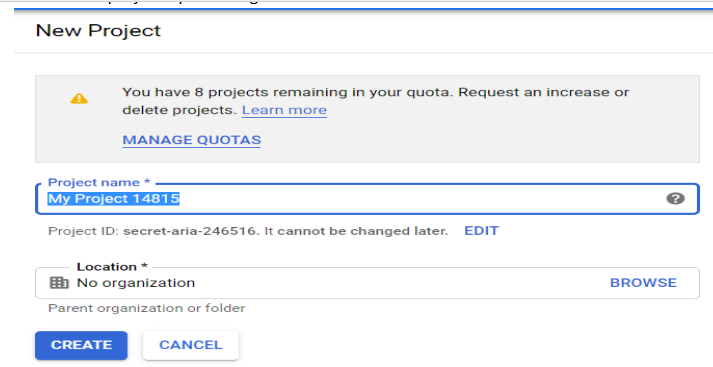
# Steps to run the pipeline

1) Login to GCP console.

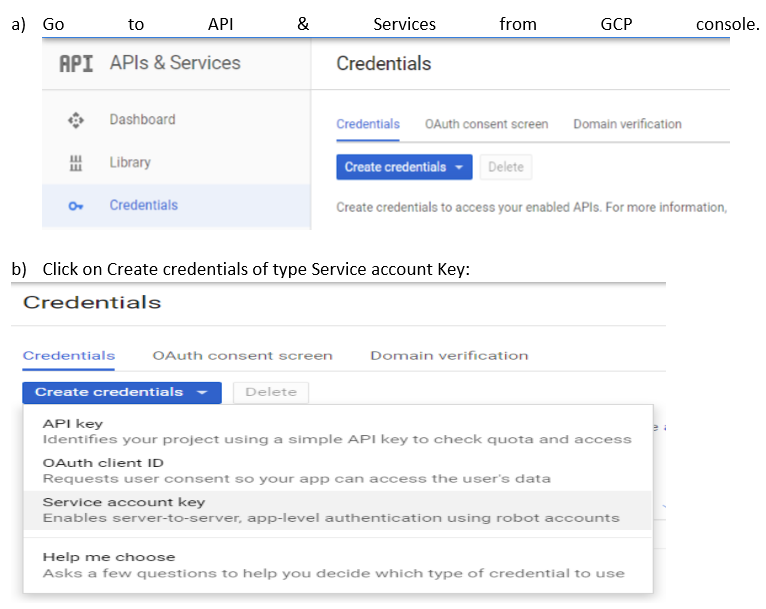
2) Select Top Left Drop Down and click on New Project.

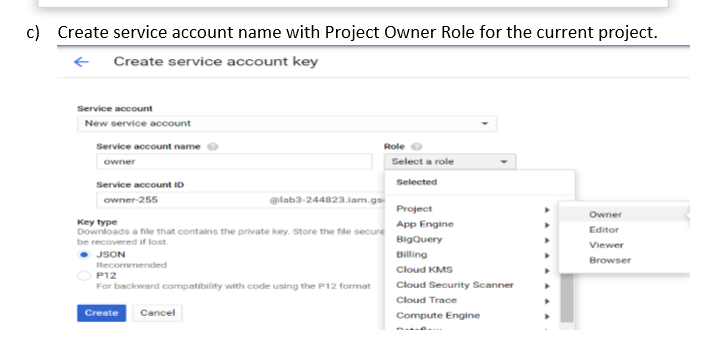


1. Create new project by selecting “NEW PROJECT” as below:



1. Create a credential file to access the Project from API

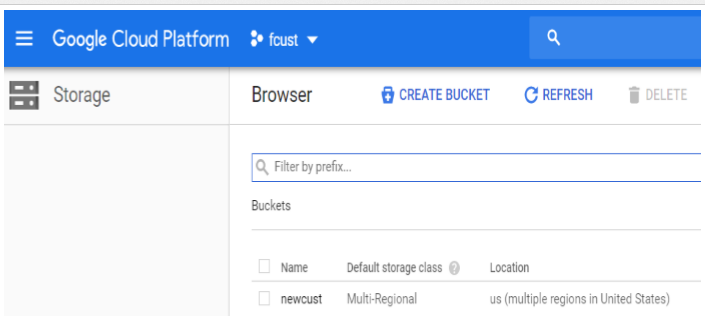




5) Next step is to create a Cloud Storage Bucket.

a) Go to Cloud Storage.

b) Create a Bucket.



1. Next step is to create Pub/Sub topic, publish message and subscription.
2. Go to Pub/Sub.
3. Create a PubSub topic. Provide a topic in the Name section.
4. Create a PubSub subscription.
5. Provide the Subscription name and copy the Topic ID from Topics and paste it in Topics ID. This will create the subscription for the topic created
6. Keep the Delivery Type as Pull and then Create the subscription.

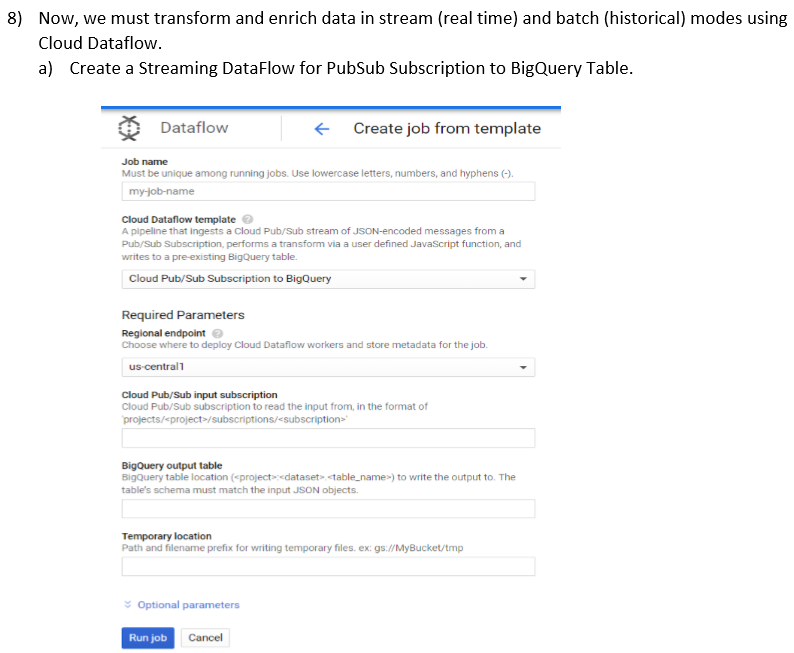
Repeat step 6 for all the data files and create a separate topic and subscription for all data files.

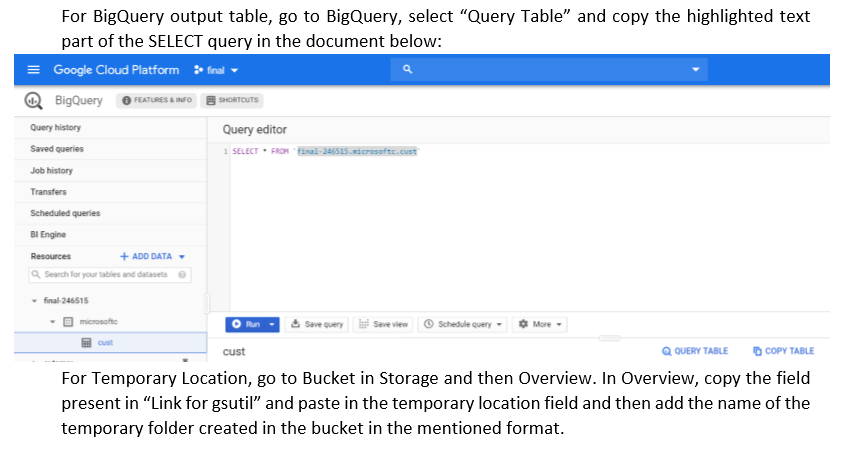
7) Go to BigQuery and create a Bigquery Table as follows.

a) Create a BigQuery dataset.

b) Select the dataset and create a BigQuery Table with the required schema under the created dataset.

c) Create tables for all the data files uploaded in storage.





Similarly, as above step, assign a Job name in the Job name text field. Choose “Text files on Cloud Storage to Cloud Pub/Sub”. In Input Cloud Storage, provide the path of the bucket where data file is saved in storage bucket overview and with that add the name of the data file in the specified format. In Output Pub/Sub Topic, copy the topic name from Topics in Pub/sub and paste it the the given field. In temporary location, paste the location of the tmp folder created in bucket in the specified format. Once all the above steps are completed, Run job.

Check if the job status is succeeded, Logs and Resource metrics for both the above-mentioned steps and check if any error is present.

Once all the above steps are successfully executed, Go to BigQuery and run the SELECT query. The data in the table should be loaded in BigQuery.

Repeat the above step for all the data files.

9) Create new VM instance on Google Compute Engine

9) Create Python Anaconda environment on the VM Intance.

9) Once we get the values in Big Query, read the data in Google Cloud compute engine using the sql query and get access to the data.

calendar="""SELECT \* FROM [mineral-aegis-248102.Project1.calendar]"""

df\_calendar = gbq.read\_gbq(salesorder,project\_id="mineral-aegis-248102",dialect="legacy")

10)Used Dask for automation.

11) Write the backend code in Flask and deploy it in the Compute Engine server.

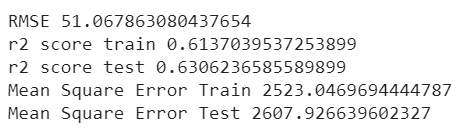
13) Implement frontend using HTML, CSS and Angular 8.

14) Deploy the code in Heroku.

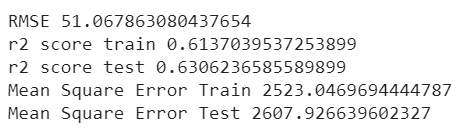
15) Create account in Google Analytics and with the link obtained by Heroku after deployment, get the link and enter in the google analytics which will further help in real time analytics of the site.

**Models Used:**

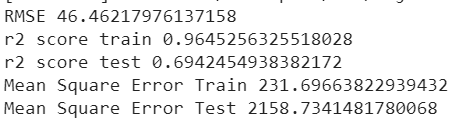
1. **Price prediction for lister:**
2. Linear Regression:

****

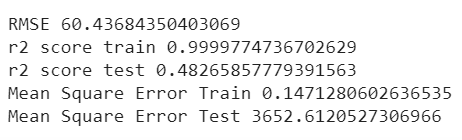
1. Random Forest Regressor:



1. XG Boost:



1. Decision Tree Regressor



Among all the models, minimum RMSE is of XGB model.

1. **Content based Recommendation:**
2. Cleaned the dataset.
3. Used term frequency – inverse document frequency to convert them to vectors.
4. Used cosine\_similarities to find the distance between them.
5. Saved the top 100 similarities.
6. Sorted the similarities according to their price for budget user.

#### Steps to Run the Project

#### Access the link below in chrome- <https://air-bnb-deploy.herokuapp.com/>

#### Fill the details required for the new listing and click on the submit button

#### It will display the predicted price for the new listing

#### Next fill in details for user to search for the hotels/homestay within the specified zip code

#### Choose if the user is on luxury travel by selecting Excursionist or choose Budget travel for restricted budget traveler.

#### 

#### In order to access the link, open the link in Google Chrome and turn off Insecure content.

#### 

#### Dashboard-

#### [*Dashboard in POWER BI*](https://app.powerbi.com/groups/me/reports/99ac1377-08bb-4d51-a5b4-7f46bc89ca69?ctid=6abfc73f-da64-4137-9f9f-15faae56f685)

#### Conclusion –

#### The Current Revenue for the year is $98299455. With the increase new listings we can target increased room from 2million to 2.2 million rooms per year. This will create an influx of additional $59million to $74million in revenue. Overall a growth of 24% can be achieved if consider the occupancy rate at 50%. Below is the summary-

|  |  |
| --- | --- |
| Current Revenue | $98299455 |
| Earlier Capacity | 1048576 |
| Increased Additional Room 5% | 1101005 |
| Increased Additional Room 15% | 1205862 |
| Occupancy rate 40% | 440402 |
| Occupancy rate 50% | 550502 |
| Revenue Increased at 40% Occupancy | $59454259 |
| Revenue Increased at 50% Occupancy | $74317824 |
|  |  |
| Growth In Revenue % | 24.40% |

#### Contribution

#### Citations:

* <https://www.kaggle.com/airbnb/seattle>
* <https://seaborn.pydata.org/>
* <http://www.numpy.org/>
* <https://jakevdp.github.io/PythonDataScienceHandbook/04.14-visualization-with-seaborn.html>
* <https://matplotlib.org/api/_as_gen/matplotlib.pyplot.html>
* <https://pandas.pydata.org/>
* <https://docs.scipy.org/doc/scipy/reference/>
* <https://scikit-learn.org/stable/documentation.html>