VIEHGROUP

Low Level Document (LLD)
Google Analytics Customer Revenue Prediction

Version number: 1.0
Last date of revision: 2 March 2023

Sugaiel Fathima A

MD Zaid N

VIEH GROUP

DECLARATION

We declare that this written submission represents us ideas is our own words and where others' ideas or words have been included, we have adequately cited and referenced the original sources.
We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in my submission.
We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

VIEHGROUP

Revision History

Version	Date	Author	Reviewer	Approver	Comments
0.1	26-02-2023	MD Zaid N	Sugaiel Fathima A		Draft version
0.2	27-02-2023	Sugaiel Fathima A	MD Zaid N		Suggested some selections like key notes, screen validations and attributes to be added
0.3	28-02-20223	MD Zaid N	Sugaiel Fathima A		Suggested document format related comments like correction of version, adding one sections for open issues etc
0.4	01-03-2023	Sugaiel Fathima A	MD Zaid N		Suggested some changes like correct sequence diagram, changes in data design sections etc
1.0	02-03-2023	MD Zaid N	Sugaiel Fathima A		Baseline Version

VIEH GROUP

Table of Contents

VIEHGROUP

1. Introduction:

1.1 Scope of the Document

- This section will cover details regarding scope of the document
- Low level design document will be at component level i.e., for website portal there will be one LLD.

1.2 Intended Audience

- This section will cover categories of audiences who will be referring/reviewing this document

1.3 System Overview

- This section will capture overview of system application i.e for what system is being developed
- Who are the stake holders of system?
- What are other external Systems through which this will be interacting

VIEHGROUP

2. Project Briefing:

In this project, we developed a predictive model using G-Store dataset to predict the total revenue for the customer that helps in better user of marketing budget and also interpret the most impacting element on the total revenue prediction using different model. As for every business, the number of customer that generate revenue will be far less than the total customers that the business interacts with. So for every business it's really important to understand, analyze and predict the areas of its revenue generation.

The main objective of this project is to create a model that predict Customer Revenue. Customer analytics is a process by which data from customer behavior is used to help make key business decisions via market segmentation and predictive analytics. This information is used by businesses for direct marketing, site selection, and customer relationship management. Marketing provides services in order to satisfy customers. With that in mind, the productive system is considered from its beginning at the production level, to the end of the cycle at the consumer. Customer analytics plays an important role in the prediction of customer behavior.

The scope of the project is Merchandise stores and other businesses can use the provided datasets to predict customer revenue. To develop this model we used Window 10 Operating system ,Visual studio software ,Project integration idea from Kaggle website.

The walkthrough of this model is as follows. Collecting history of customer data from the merchandise store , then predicting customer revenue from the datasets and then the final prediction.

VIEHGROUP

3. Problem Statement:

Predict the total revenue per customer to establish a better marketing Decison.

4. Problem Solution:

Develop a predictive model using G-Store dataset to predict the total revenue for the customer that helps in better user of marketing budget and also interpret the most impacting element on the total revenue prediction using different model.

5. Objective of the Project:

Objective of this project is to create a model that predict Customer Revenue.

6. Scope of Project:

Merchandise stores and other businesses can use the provided datasets to predict customer revenue.

7. Block Diagram:

Collecting history of customer data
from the merchandise store

predicting customer revenue
from the datasets

Final
Prediction

VIEHGROUP

8. Requirements Gathering:

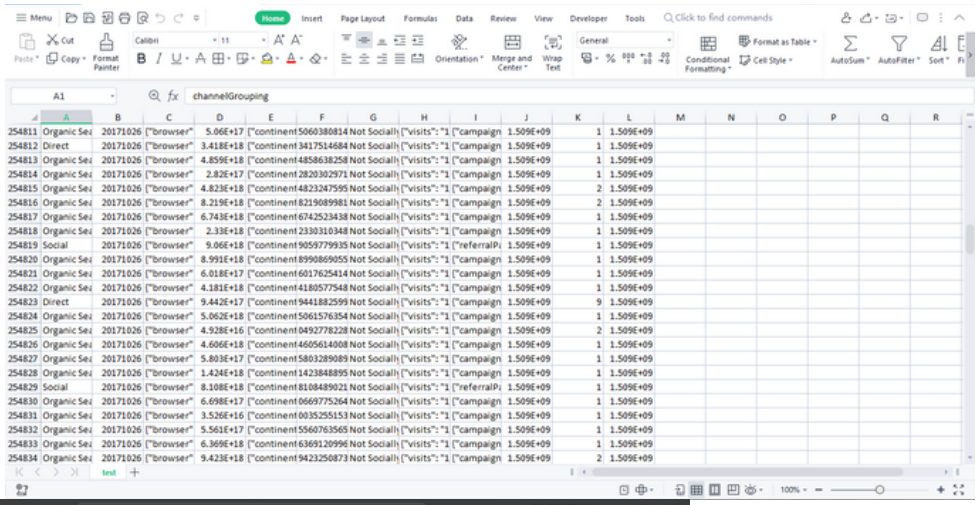
- Window 10 Operating system
- Visual studio software
- 2 Team members for the research part
- Project integration idea from Kaggle website
- Few Github Non copyrighted source codes

9. Analysis:

In this project we develop a predictive model using G-Store dataset to predict the total revenue for the customer that helps in better user of marketing budget and also interpret the most impacting element on the total revenue prediction using different model. Google Analytics automatically enriches customer's data by bringing Google machine-learning expertise to bear on customer's dataset to predict the future behavior of the customer. As for every business, the number of customer that generate revenue will be far less than the total customers that the business interacts with. So for every business it's really important to understand, analyze and predict the areas of its revenue generation. In the first step, we collect history of customer data from the merchandise store. We store all the collected data in the datasets. In the second step, we try to predict the customer revenue from the collected datasets. So objective of this project is to create a model that predict Customer Revenue. In the third step, we will derive the final prediction the will be depicted in the form of graphs, and the graphs are knowns as Bar charts.

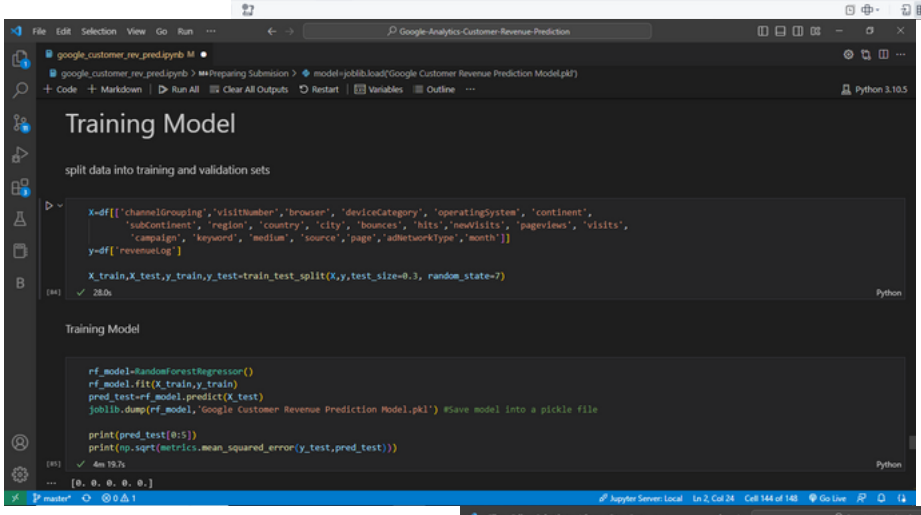
VIEHGROUP

10. Final Screenshot of Project Output



channelGrouping

	channelGrouping	date	device	totals	trafficSource
254811	Organic Sea	20171026	{"browser"	1.509E+09	1.509E+09
254812	Direct	20171026	{"browser"	1.509E+09	1.509E+09
254813	Organic Sea	20171026	{"browser"	1.509E+09	1.509E+09
254814	Organic Sea	20171026	{"browser"	1.509E+09	1.509E+09
254815	Organic Sea	20171026	{"browser"	1.509E+09	1.509E+09
254816	Organic Sea	20171026	{"browser"	1.509E+09	1.509E+09
254817	Organic Sea	20171026	{"browser"	1.509E+09	1.509E+09
254818	Organic Sea	20171026	{"browser"	1.509E+09	1.509E+09
254819	Social	20171026	{"browser"	1.509E+09	1.509E+09
254820	Organic Sea	20171026	{"browser"	1.509E+09	1.509E+09
254821	Organic Sea	20171026	{"browser"	1.509E+09	1.509E+09
254822	Organic Sea	20171026	{"browser"	1.509E+09	1.509E+09
254823	Direct	20171026	{"browser"	1.509E+09	1.509E+09
254824	Organic Sea	20171026	{"browser"	1.509E+09	1.509E+09
254825	Organic Sea	20171026	{"browser"	1.509E+09	1.509E+09
254826	Organic Sea	20171026	{"browser"	1.509E+09	1.509E+09
254827	Organic Sea	20171026	{"browser"	1.509E+09	1.509E+09
254828	Organic Sea	20171026	{"browser"	1.509E+09	1.509E+09
254829	Social	20171026	{"browser"	1.509E+09	1.509E+09
254830	Organic Sea	20171026	{"browser"	1.509E+09	1.509E+09
254831	Organic Sea	20171026	{"browser"	1.509E+09	1.509E+09
254832	Organic Sea	20171026	{"browser"	1.509E+09	1.509E+09
254833	Organic Sea	20171026	{"browser"	1.509E+09	1.509E+09
254834	Organic Sea	20171026	{"browser"	1.509E+09	1.509E+09



Training Model

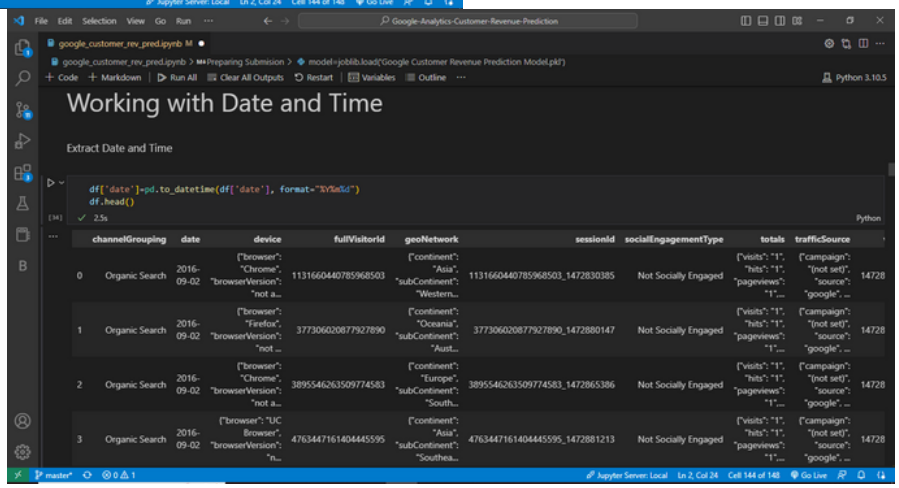
split data into training and validation sets

```
X=df[['channelGrouping','visitNumber','browser', 'deviceCategory', 'operatingSystem', 'continent',
      'subContinent', 'region', 'country', 'city', 'bounces', 'hits','newVisits', 'pageviews', 'visits',
       'campaign', 'keyword', 'medium', 'source','page','adNetworkType','month']]
y=df['revenueLog']

X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.3, random_state=7)
```

Training Model

```
rf_model=RandomForestRegressor()
rf_model.fit(X_train,y_train)
pred_test=rf_model.predict(X_test)
joblib.dump(rf_model,'Google Customer Revenue Prediction Model.pkl') #Save model into a pickle file

print(pred_test[0:5])
print(np.sqrt(metrics.mean_squared_error(y_test,pred_test)))
```

```
[0. 0. 0. 0. 0.]
```



Working with Date and Time

Extract Date and Time

```
df['date']=pd.to_datetime(df['date'], format="%Y%m%d")
df.head()
```

	channelGrouping	date	device	fullVisitorId	geoNetwork	sessionId	socialEngagementType	totals	trafficSource
0	Organic Search	2016-09-02	{"browser": "Chrome", "browserVersion": "not a...	1131660440785968503	{"continent": "Asia", "subContinent": "Western...	1131660440785968503_1472830385	Not Socially Engaged	{"visits": "1", "hits": "1", "pageviews": "1",...	{"campaign": "(not set)", "source": "google", ... 14728
1	Organic Search	2016-09-02	{"browser": "Firefox", "browserVersion": "not ...	377306020877927890	{"continent": "Oceania", "subContinent": "Aust...	377306020877927890_1472880147	Not Socially Engaged	{"visits": "1", "hits": "1", "pageviews": "1",...	{"campaign": "(not set)", "source": "google", ... 14728
2	Organic Search	2016-09-02	{"browser": "Chrome", "browserVersion": "not a...	3895546263509774583	{"continent": "Europe", "subContinent": "South...	3895546263509774583_1472865386	Not Socially Engaged	{"visits": "1", "hits": "1", "pageviews": "1",...	{"campaign": "(not set)", "source": "google", ... 14728
3	Organic Search	2016-09-02	{"browser": "UC Browser", "browserVersion": "n...	4763447161404445595	{"continent": "Asia", "subContinent": "Southea...	4763447161404445595_1472881213	Not Socially Engaged	{"visits": "1", "hits": "1", "pageviews": "1",...	{"campaign": "(not set)", "source": "google", ... 14728