GreedyGame Data Analyst Assignment Solutions

Introduction: -

- The dataset is about the apps which contains information like App, Ad Group, State, Revenue, etc.
- There are 48293 rows and 12 Columns
- Also, it consists of 4016 missing values in the dataset. Only 8% of data is missing from entire data
- Dropping these missing values from the dataset and analysing the data further

1. **How would you increase the overall revenue for an app?**

```
1 df.corr()
```

	Revenue	Ad Requests	Ad Responses	Impressions	1st Clicks	2nd Clicks
Revenue	1.000000	0.127370	0.220613	0.192691	0.332629	0.329001
Ad Requests	0.127370	1.000000	0.330752	0.321995	0.193078	0.238182
Ad Responses	0.220613	0.330752	1.000000	0.996202	0.435179	0.400689
Impressions	0.192691	0.321995	0.996202	1.000000	0.416792	0.376401
1st Clicks	0.332629	0.193078	0.435179	0.416792	1.000000	0.634154
2nd Clicks	0.329001	0.238182	0.400689	0.376401	0.634154	1.000000

i. From Entire Table, there's a high correlation between Ad Responses and Impressions.
ii. In terms of Revenue, there's a high correlation b/w Revenue & 1st Clicks followed by 2nd Clicks.
iii. Also, if we see 1st Clicks & 2nd clicks have 63% correlation meaning users tend to interact more after 1st clicks to know more about an ad/ content.
iv. To increase overall revenue of the apps, we can Raise the prices or increase our customer base.
v. Around 87% of data isn't generating revenue. So, making more ad content on those specific apps can improve Revenue

2. **Are there any discrepancies in the data set shared with you? What are the possible reasons for such mismatch?**

A. From data information,

Fill Rate--> % of Ad Responses on Ad Requests

Render Rate--> % of Impressions on Ad Responses

There are few discrepancies or fault information in column fill rate

13	06 Sep 2020	Callbreak Multiplayer	Rooter_Callbreak	Uttar Pradesh	0.52	3702040	190575	100%	177858	93.3%	1064	1049
14	06 Sep 2020	Callbreak Multiplayer	Callbreak Multiplayer2_Inmobi_No Floor	Uttar Pradesh	1.37	3702040	18797	100%	14648	77.93%	384	120

```
###Fill Rate--> % of Ad Responses on Ad Requests
###Render Rate--> % of Impressions on Ad Responses
```

There are few discrepancies or fault information in column Fill Rate

```
#### For Instance, Lets look for few rows.
#index- 13
190575/3702040
```

```
0.05147837408563927
```

```
#index-14
18797/3702040
```

```
0.005077470799883308
```

- As we can see that, there's a mismatch in fill rate column for index 13 & 14.
- I also randomly checked few other rows, and found the same mismatch except few rows
- We can change this calculation and move ahead.

```
df.isnull().sum()
```

```
Date              0
App              23
Ad Group       3993
State             0
Revenue           0
Ad Requests       0
Ad Responses      0
Fill Rate         0
Impressions       0
Render Rate       0
1st Clicks        0
2nd Clicks        0
dtype: int64
```

- Also, the dataset has 4016 missing values which can make a difference while analysing data
- Only App & Ad Group have missing values but those are the important columns in dataset for analysing data.
- So, I just dropped them from the dataset without replacing them with mean or mode as it may not be the right approach always

3. **If you had a classification mechanism, which apps would you classify as Tier 1 or Important category for GG? Explain your reasoning.**
    - Around 86% of data has zero Revenue. So, I have come up with a strategy to filter data with mean value of the revenue. The Mean value of the Revenue column from original data is 0.07844
    - So, I have filtered the data with name **Revenue_mean** which has 2566 Rows & 12 Columns
    - Then, I found out the mean value of Revenue_mean which is 1.322404 greater than original dataset
    - In Revenue_mean, we have 30 Unique Apps and has highest revenue of US $42.39
    - The Total Sum of Revenue in Revenue_mean is US $3393.29
    - While Total Sum of Revenue in Original Dataset is US $3473.21
    - So, we can say that above 30 Unique Apps has US $3393.29 while the other apps have just US $79.92
    - So, I can say these specific 30 Apps can be categorized has Tier 1 or Important Apps for GG and rest has Tier-2

4. **Are there any ad groups which are causing an opportunity loss for the app? If yes, how would you mitigate it?**

A. I tried to answer this but I failed to figure out.

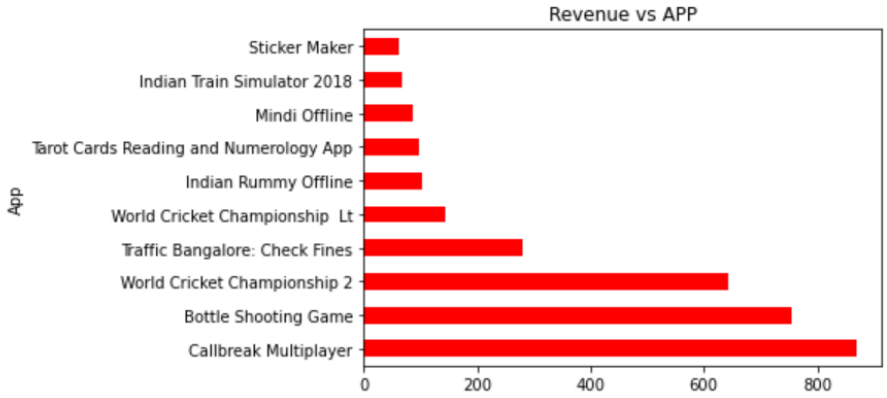
Additional Insights: -

1. Top 10 Apps Based on Revenue

1. Top 10 Apps based on Revenue

```
df.groupby(['App'])['Revenue'].sum().sort_values(ascending = False)[:10].plot(kind = "barh", tit
```

```
<AxesSubplot:title={'center':'Revenue vs APP'}, ylabel='App'>
```

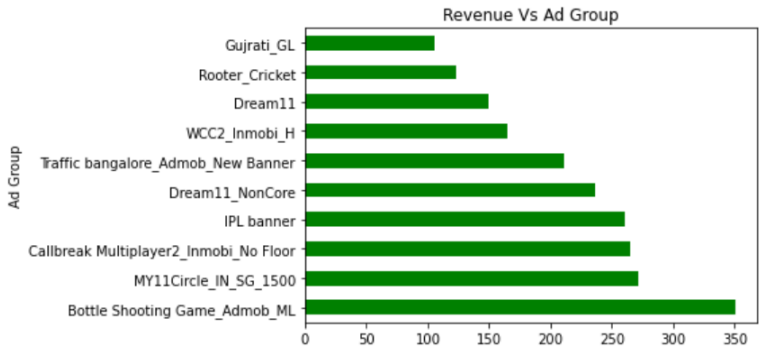


- CallBreak Multiplayer App has made the highest Revenue
- These are the top 10 Apps wrt Revenue. 7 out of these 10 Apps, were Games
- Rest of the apps are to check fines (Traffic Bangalore), Numerology App, & Sticker Maker

2. Top 10 Ad Groups based on Revenue

```
df.groupby(['Ad Group'])['Revenue'].sum().sort_values(ascending = False)[:10].plot(kind = "barh", color = "gree
```

```
<AxesSubplot:title={'center':'Revenue Vs Ad Group'}, ylabel='Ad Group'>
```



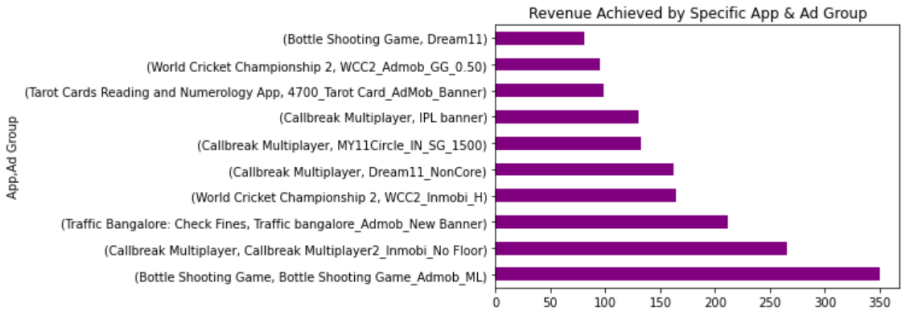
- Ad Group "Bottle Shooting Game_admob_ML" has highest revenue of 350 followed by My11Circle, etc

i. Revenue Achieved by Specific App & Ad Group

Revenue Achieved by Specific App & Ad Group

```
1 _values(ascending = False)[:10].plot(kind = "barh", color = "purple", title = "Revenue Achieved by Specific App & Ad Grou
```

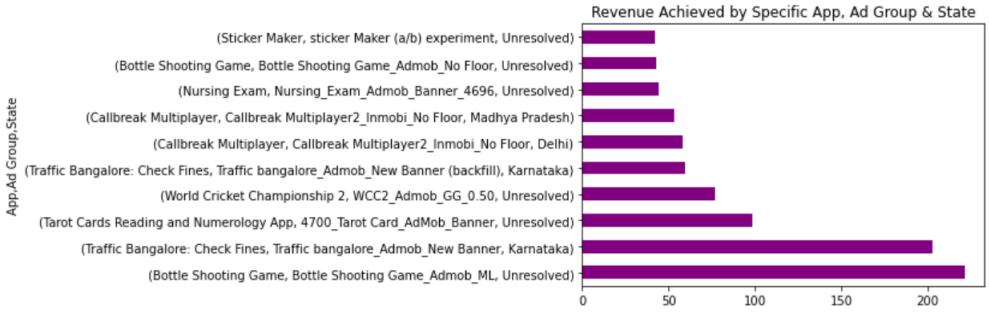
```
<AxesSubplot:title={'center':'Revenue Achieved by Specific App & Ad Group'}, ylabel='App,Ad Group'>
```



3. Revenue Achieved by Specific App, Ad Group State-wise

```
1 ort_values(ascending = False)[:10].plot(kind = "barh", color = "purple", title = "Revenue Achieved by Specific App, Ad Group &
```

```
<AxesSubplot:title={'center':'Revenue Achieved by Specific App, Ad Group & State'}, ylabel='App,Ad Group,State'>
```



- Bottle Shooting Game has achieved overall highest revenue of 288.97 interestingly it's not from India
- Traffic Bangalore App which is from Karnataka has achieved highest revenue of US $262.52 in India.
- These Top 10 Revenue Achieved by Apps state-wise has mostly Gaming Apps, a Traffic App and Numerology App.
- Tells, how gaming industry is booming worldwide.

4. 1st Clicks, 2nd Clicks based on Apps (Plot in EDA_Jupyter Notebook)

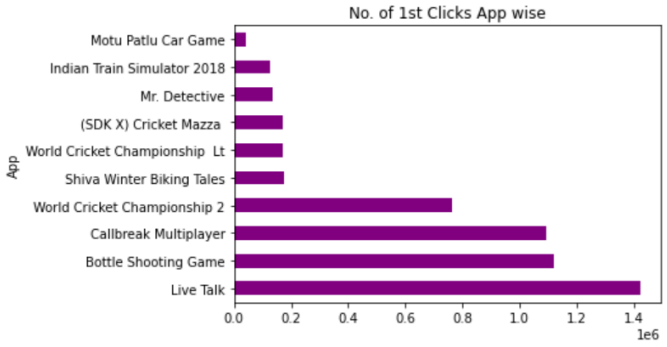
A. Number of 1st Clicks App wise

- Live Talk App has most no. of 1st Clicks followed by Bottle Shooting Game, CallBreak multiplayer, and so on

A. Number of 1st Clicks App wise

```
df.groupby(['App'])['1st Clicks'].sum().sort_values(ascending = False)[:10].plot(kind = "barh", color = "purple", titl
```

```
<AxesSubplot:title={'center':'No. of 1st Clicks App wise'}, ylabel='App'>
```

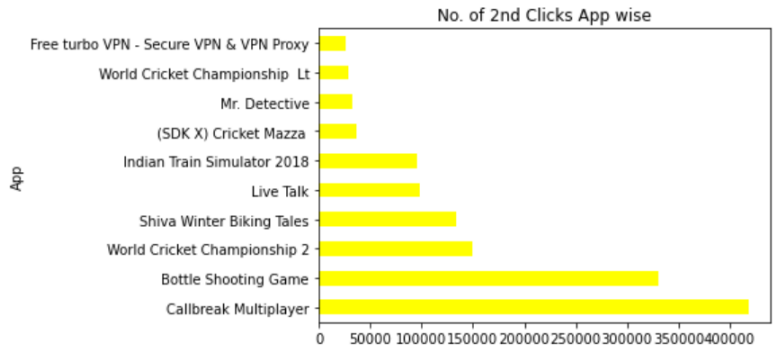


B. Number of 2nd Clicks App-wise

- CallBreak Multiplayer has highest no. of 2nd Clicks followed by Bottle Shooting Game and So on.

```
df.groupby(['App'])['2nd Clicks'].sum().sort_values(ascending = False)[:10].plot(kind = "barh", color = "yellow
```

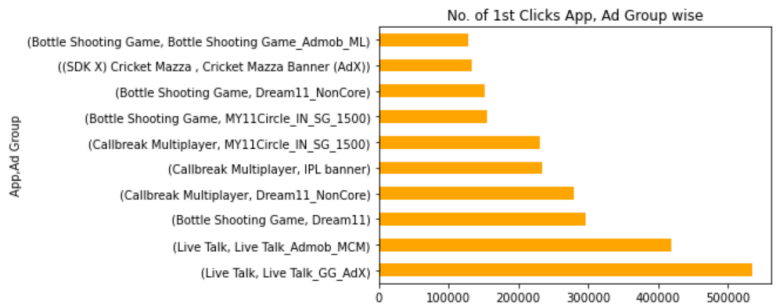
```
<AxesSubplot:title={'center':'No. of 2nd Clicks App wise'}, ylabel='App'>
```



C. Number of 1st Clicks based on App & Ad Group

```
df.groupby(['App', 'Ad Group'])['1st Clicks'].sum().sort_values(ascending = False)[:10].plot(kind = "barh", color = "orange"
```

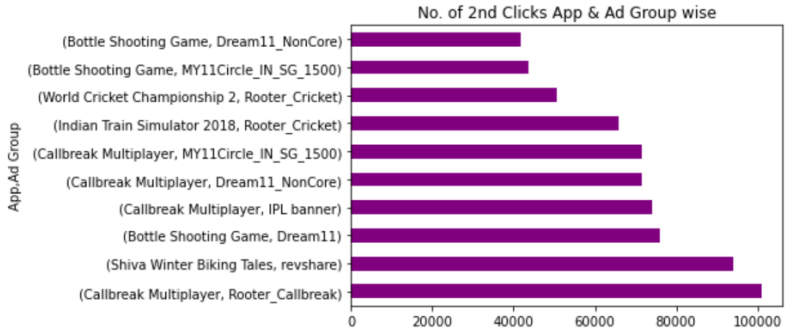
```
<AxesSubplot:title={'center':'No. of 1st Clicks App, Ad Group wise'}, ylabel='App,Ad Group'>
```



D. Number of 2nd Clicks based on App & Ad Group

```
df.groupby(['App', 'Ad Group'])['2nd Clicks'].sum().sort_values(ascending = False)[:10].plot(kind = "barh", color =
```

```
<AxesSubplot:title={'center':'No. of 2nd Clicks App & Ad Group wise'}, ylabel='App,Ad Group'>
```



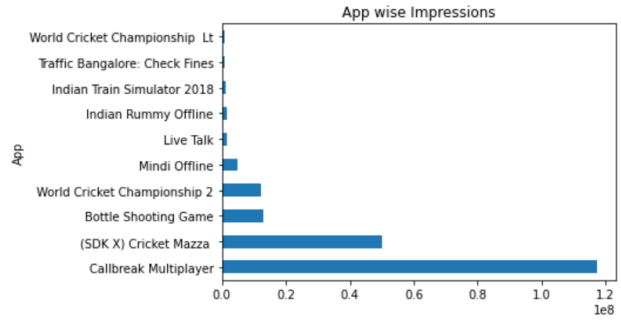
- In Summary related to 1st Clicks & 2nd Clicks, we can say most of the Clicks were of Games and Live Chat Apps. Except few are other non-gaming apps

5. Impressions

A. Top 10 Impressions App wise

```
df.groupby(['App'])['Impressions'].sum().sort_values(ascending = False)[:10].plot(kind = "barh", title = "App wise Impressior
```

```
<AxesSubplot:title={'center':'App wise Impressions'}, ylabel='App'>
```

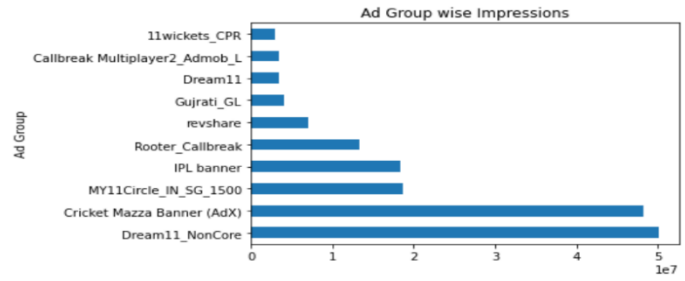


- CallBreak Multiplayer Application has highest number of Impressions followed by Cricket Mazza

B. Top 10 Impressions Ad Group wise

```
df.groupby(['Ad Group'])['Impressions'].sum().sort_values(ascending = False)[:10].plot(kind = "barh", title = "Ad Gro
```

```
<AxesSubplot:title={'center':'Ad Group wise Impressions'}, ylabel='Ad Group'>
```

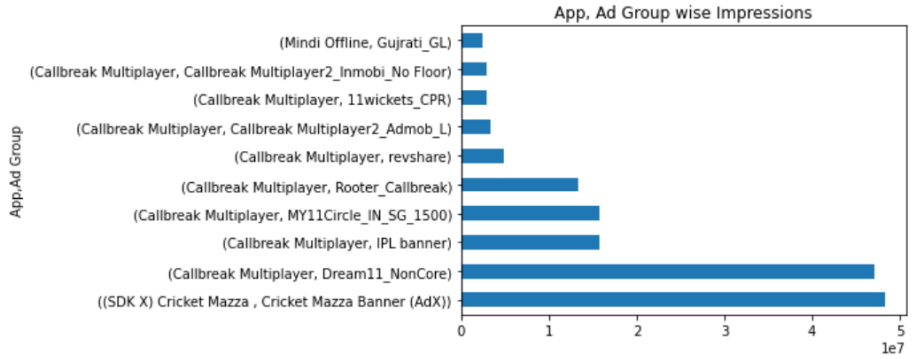


- Dream11_noncore Ad has highest Impressions followed by Cricket Mazza and So on

C. Top 10 Impressions App & Ad Group wise

```
df.groupby(['App', 'Ad Group'])['Impressions'].sum().sort_values(ascending = False)[:10].plot(kind = "barh", tit
```

```
<AxesSubplot:title={'center':'App, Ad Group wise Impressions'}, ylabel='App,Ad Group'>
```



- Here, if we can see Cricket Mazza App has highest Impressions followed by CallBreak Multiplayer in terms of App, Ad Group
- Interesting Insight, here is Dream11_NonCore, IPL Banner, My11 Circle Ads, 11wickets all are Fantasy Games appearing CallBreak Multiplayer which is a card game. These Ads are appearing on user's screen to Bet or use these Apps to get new users.

6. Ad Requests

A. No. of Ad Requests made by App State-wise

A. No. of Ad Requests made by App Statewise

```
df.groupby(['State', 'App'])['Ad Requests'].sum().sort_values(ascending = False)[:10].plot(kind = "barh", t
```

```
<AxesSubplot:title={'center':'No. of Ad Requests by an App Statewise'}, ylabel='State,App'>
```

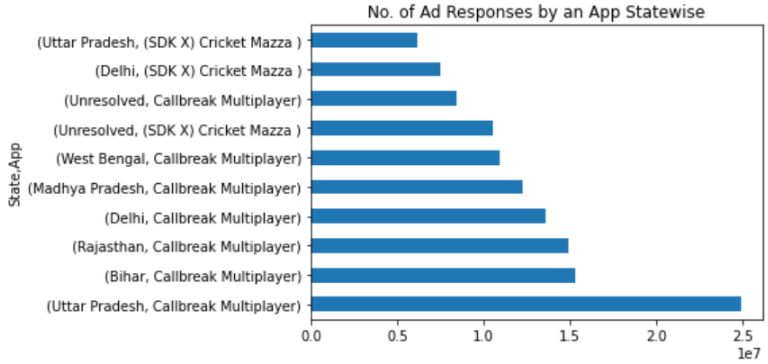
- Callbreak Multiplayer App has requested most no. of Ad Requests from state UP followed by Bihar and so on.
- Amongst these top 10 Apps, Majority of Ad Requests has made by Callbreak Multiplayer App from most of Northern & Eastern Part of India

B. No. of Ad Responses made by App State-wise

B. No. of Ad Responses made by App Statewise

```
1  df.groupby(['State', 'App'])['Ad Responses'].sum().sort_values(ascending = False)[:10].plot(kind = "barh", tit
```

```
<AxesSubplot:title={'center':'No. of Ad Responses by an App Statewise'}, ylabel='State,App'>
```

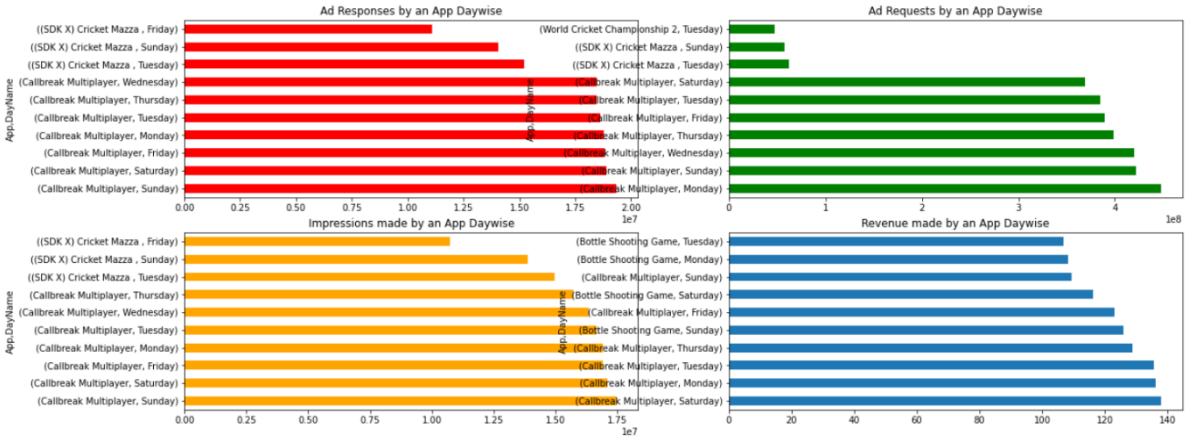


- After requesting for an Ad, Most No. of Ad Responses is from Callbreak Multiplayer of state UP

7. Analysing Dates

- The dataset is from 06th Sept 2020 to 12th Sept 2020
- It's a period of 7 days
- Extracted Day and Week name from Date for further analysis

A. Revenue, Ad Requests, Ad Responses, Impressions made by App Day-wise



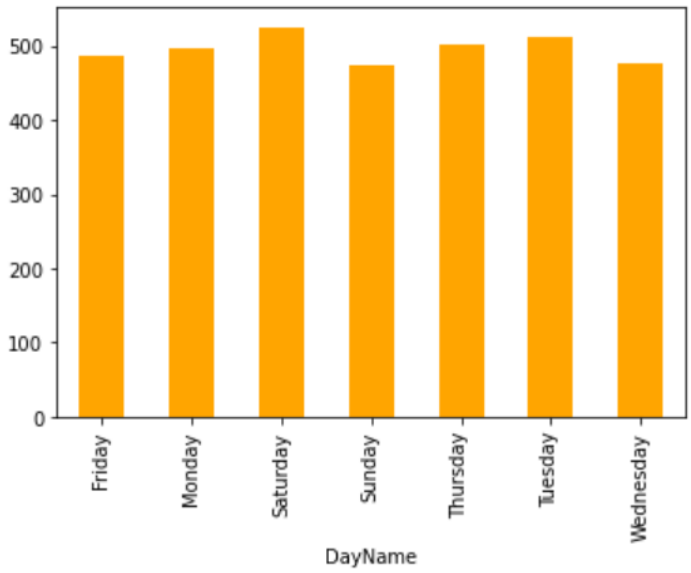
- Callbreak Multiplayer game has outplayed amongst other Apps and above plot shows the Revenue, Impressions, Ad Responses, Ad Requests Day-wise
- Highest Revenue is achieved by Callbreak Multiplayer on Saturday.

B. Revenue Achieved Day-wise

Revenue Day-wise ¶

```
1  df.groupby(['DayName'])['Revenue'].sum().plot(kind = "bar", color = "orange")
```

```
<AxesSubplot:xlabel='DayName'>
```



- Saturday has achieved highest revenue of US $525.14

GitHub Link for Notebook

https://github.com/siddhu21/Sample