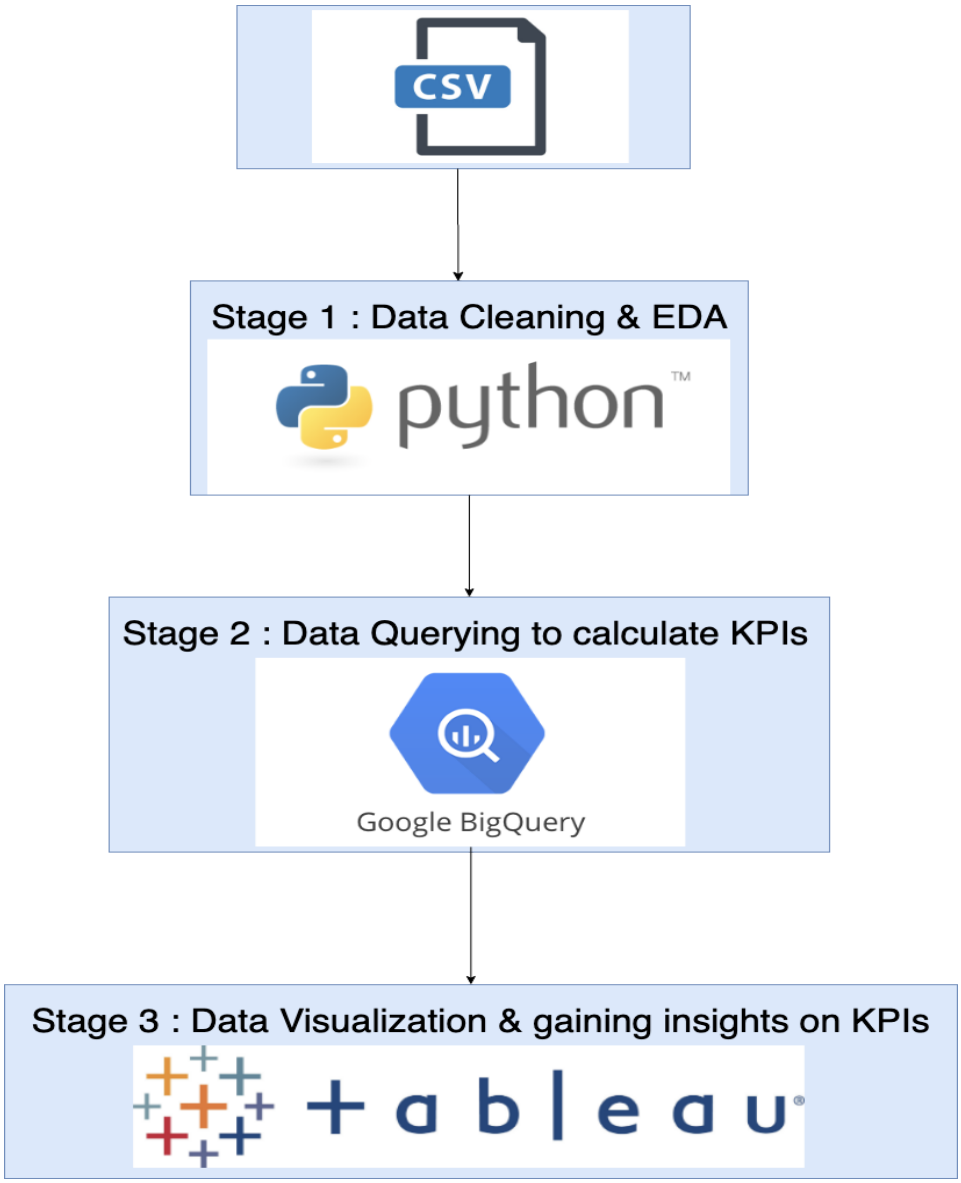
Qatch Analytics Project Writeup

- Rohan Kapadnis

---

1. Architecture of the Project (Data Pipeline)



- The architecture of this project is shown above. I have tried to create a data pipeline where one component is connected to another for reliability and consistency.
- After reading the data in Jupyter Notebook, I have performed data cleaning(explained in next segment)
- Next segment is about migrating the three dataframes (outbound_text, reactions, purchases) in Google BigQuery. To do that I have used the **pandas_gbq** module.

```
In [124]: pandas_gbq.to_gbq(dataframe, destination_table="qatch.userInfo", project_id = "qatch-technical-project")
```

-
- Above is the code snippet
- Next step is to query results using BigQuery from the data we just loaded
- After performing and calculating the KPIs, I connected BigQuery to Tableau for visualization over the data
- This connection is made possible by the built in option to connect server in Tableau under the “to a server” tab. After going into the tab, select Google BigQuery and then enter the google credentials.
- Finally the pipeline is connected.
- I then performed visualization of the KPIs and gain insight and in the process answer the question of : How can we measure success?

2. The Data Cleaning stage

- The data is fairly clean. There were no null values!
- There was no need for conversion of data types
- Rounded off the commission to two decimal places

Rounding the Commission to 2 decimal places

```
In [92]: purchases['commission']= purchases['commission'].round(2)
```

```
outbound_text.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 22334 entries, 0 to 22333
Data columns (total 3 columns):
 #   Column     Non-Null Count  Dtype 
---  ------     --------------  ----- 
 0   text_id    22334 non-null  int64 
 1   uid        22334 non-null  int64 
 2   text_date  22334 non-null  object
dtypes: int64(2), object(1)
memory usage: 523.6+ KB
```

```
reaction.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 7290 entries, 0 to 7289
Data columns (total 4 columns):
 #   Column               Non-Null Count  Dtype 
---  ------               --------------  ----- 
 0   inbound_text_id      7290 non-null   int64 
 1   inbound_ text _date  7290 non-null   object
 2   reaction_type        7290 non-null   object
 3   outbound_text_id     7290 non-null   int64 
dtypes: int64(2), object(2)
memory usage: 227.9+ KB
```

```
purchases.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 297 entries, 0 to 296
Data columns (total 5 columns):
 #   Column            Non-Null Count  Dtype  
---  ------            --------------  -----  
 0   purchase id       297 non-null    int64  
 1   purchase_date     297 non-null    object 
 2   amount            297 non-null    float64
 3   commission        297 non-null    float64
 4   outbound_text_id  297 non-null    int64  
dtypes: float64(2), int64(2), object(1)
memory usage: 11.7+ KB
```

3. SQL queries to generate views KPIs

1. **Number of Reactions per user**

ReactionNumbersPerUser

```
select count(a.text_id), a.uid
from qatch.outbound_text a
left join qatch.reactions b
on b.outbound_text_id =  a.text_id
where b.inbound_text_id IS NOT NULL
group by a.uid
order by a.uid
```

2. **Received to purchase ratio**

ReceivedToPurchaseRatio

```
select a.uid as UserID, count(a.text_id) as TotalReceived, count(b.purchase_id) as TotalPurchased
from
`qatch.outbound_text` a
left join
`qatch.purchases` b
on
a.text_id = b.outbound_text_id
group by UserID
order by UserID desc
```

3. **Total Amount per user**

TotalAmountPerUser

```
select a.uid, sum(b.amount) as totalAmount
from `qatch.outbound_text` a
inner join
`qatch.purchases` b
on a.text_id = b.outbound_text_id
group by a.uid
order by totalAmount desc
```

4. **Total Purchases per user**

TotalPurchasesPerUser

```
select a.uid as UserID , count(b.purchase_id) as TotalPurchases from `qatch.outbound_text` a
inner join
`qatch.purchases` b
on
a.text_id = b.outbound_text_id
group by a.uid
order by TotalPurchases desc
```

5. **Total Received vs Total Reacted**

TotalReceivedVSTotalReacted

```
SELECT T1.userID, T2.totalReceived, T1.totalReactions  from

(select count(a.text_id) AS totalReactions , a.uid as userID
from qatch.outbound_text a
left join qatch.reactions b
on b.outbound_text_id =  a.text_id
where b.inbound_text_id IS NOT NULL
group by a.uid
order by a.uid) AS T1,

(select count(a.text_id) AS totalReceived, a.uid as userID
from qatch.outbound_text a
left join qatch.reactions b
on b.outbound_text_id =  a.text_id
group by uid) AS T2


WHERE T1.userID = T2.userID
Order by T1.userID;

```

Note : There are many other KPIs which I have directly visualized in Tableau without querying.

4. KPIs - How can we measure success?

We can measure success based on Purchases, Amount, Commission earned, Reactions and User Activity. I have developed the following KPIs which explore the mentioned domains of how we can measure success.

1. Received to Reaction Rate Per User

- This KPI takes in the total number of texts one user receives and total number of reactions by the user

- The formula is given by :
  [Total Reactions] / [Total Texts Received] * 100

- This KPI helps in understanding what is the rate at which Qatch's recommendations are being reacted to. It will help in understanding whether the recommendations are relevant to the user and how engaged the users are.

- By visualizing using bar plots, the rate ranged between 40 to 100(I have only considered users who received a minimum of 60 messages so that ones with low number of receiving messages are filtered out as they have significantly high rate but the sample is less).

- Users near 100 are the ones which are actively reacting to the recommendations. They help in improving their future recommendations. But the ones which are to the lower end need to be actively engaged in order to get better recommendations.

- The company is doing well when it comes to this KPI as the average rate is around 93(data might be skewed due to outliers). But to improve the success, Qatch needs to nudge users on a regular basis whose reaction rate is low. This would prove to be helpful in terms of recommendation. It would also act as a reminder to customers who are not engaging much.

- I have used BigQuery to come up with the total reactions and total texts sent. And then applied a calculated field to find the rate for each user which was later

filtered on users who received a minimum of 60 texts to handle the problem of Sample Size Limitation.

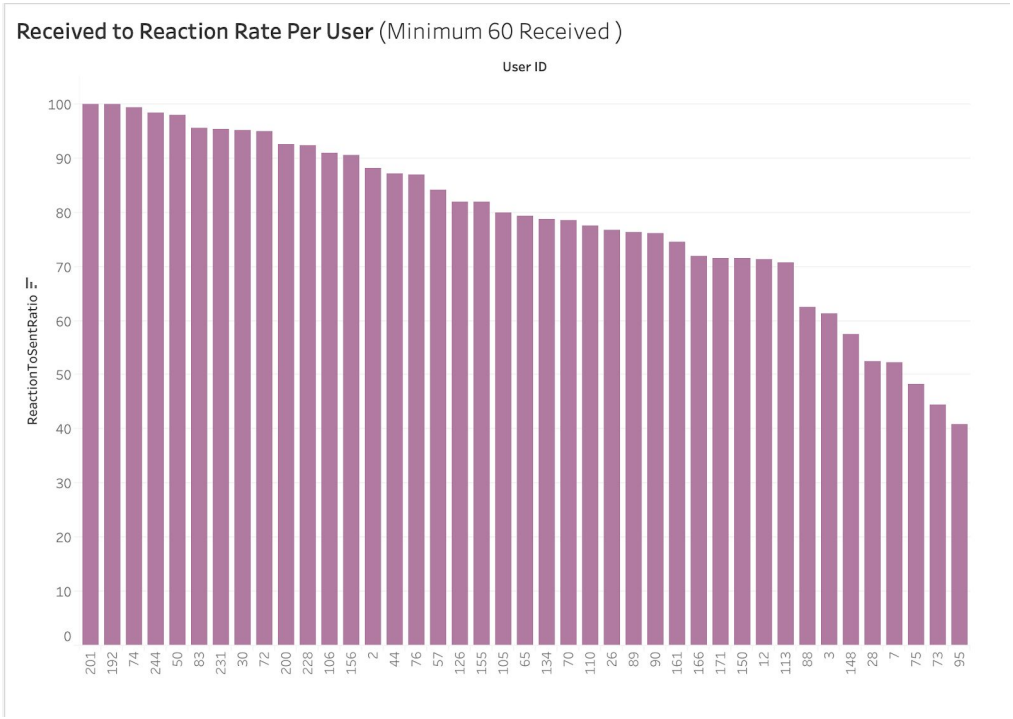
```sql
SELECT T1.userID, T2.totalReceived, T1.totalReactions  from

(select count(a.text_id) AS totalReactions , a.uid as userID
from qatch.outbound_text a
left join qatch.reactions b
on b.outbound_text_id =  a.text_id
where b.inbound_text_id IS NOT NULL
group by a.uid
order by a.uid) AS T1,

(select count(a.text_id) AS totalReceived, a.uid as userID
from qatch.outbound_text a
left join qatch.reactions b
on b.outbound_text_id =  a.text_id
group by uid) AS T2


WHERE T1.userID = T2.userID
Order by T1.userID;
```

-



-

- The above images are for quick understanding. I have also included the tableau dashboard files which can be interacted with and be used to filter out as per the UserID of liking

2. Received to Purchase Ratio Per User(Conversion Rate)

- This KPI is similar to the previous one. But instead of the number of Reactions, I have taken total number of Purchases

- The formula is given by :
  [Total Purchases] / [Total Texts Received] * 100

- This KPI helps in understanding whether the recommendations are relevant to the user and how those recommendations are getting converted to purchases by the users.

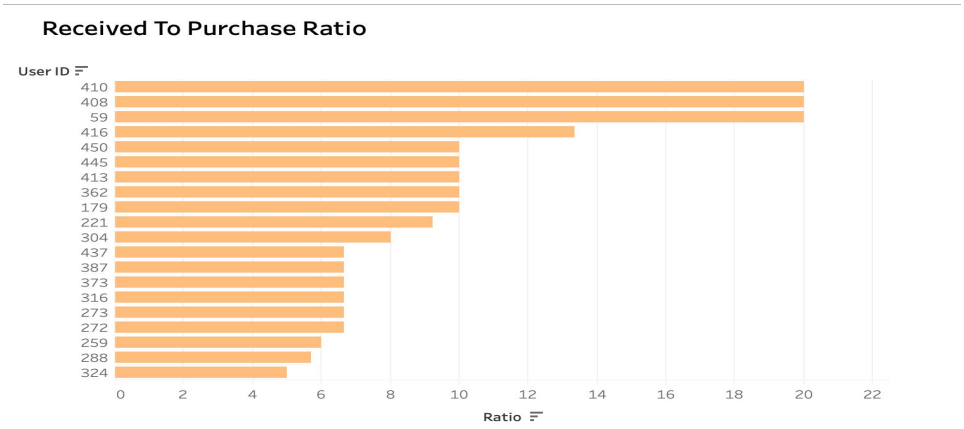
- Users with high conversion rates are getting the best recommendations. Users with low conversion rates should be dealt with and Qatch should try to improve their recommendations on those particular users.

- There is a conversion rate of 2 to 20 for the top users users w.r.t. the conversion rate.

- The industry standard for conversion rate to be “good” is 2% and beyond. So the company is currently sitting at a good spot when it comes to recommendations being converted in purchases. It can be considered success in its own form.

- But, about the customers which fall below this standard of 2%, better recommendations need to provided, some promotional offers must be given and user should be nudged from time to time if the engagement rate is low

- Information about the product category can help tremendously to understand which category is most likely being reacted to along with the user demographics information



-

```sql
select a.uid as UserID, count(a.text_id) as TotalReceived, count(b.purchase_id) as TotalPurchased
from
`qatch.outbound_text` a
left join
`qatch.purchases` b
on
a.text_id = b.outbound_text_id
group by UserID
order by UserID desc
```

3. Sales Growth

- Sales growth is a very important KPI when it comes to marketing analysis. In the end, it is the measure of a company's sales and growth which would determine how successful it is.

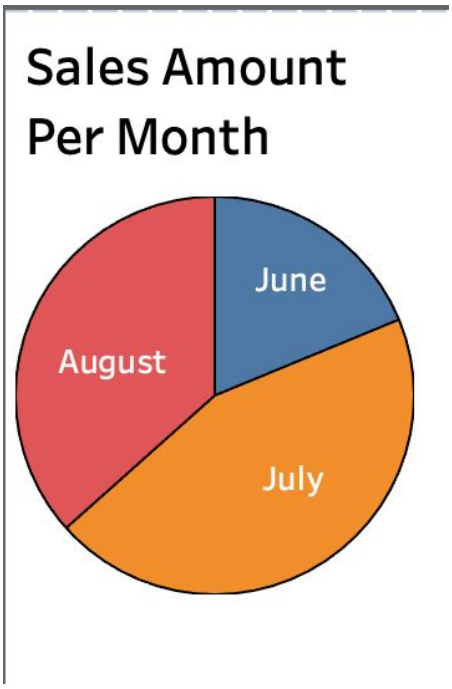
- I have measured sales growth by calculating the sum of total amount per month

- The total amount from the purchase earned increased from June to July. But it decreased from July to August. So there was an increase first and then a decrease.

- Interestingly, there was an increase in total number of purchases per month from June to July and from July to August. Hence it can be said that even though the number of purchases increased, the total amount earned did not which can be due to the purchase of products which are cheaper as compared to the ones bought in July

- So the company having an increase in sales is a good sign of success, but the decrease should not be taken as a negative because the number of purchases are increasing and customers are buying relatively cheaper products

- Hence, it is an indication that users prefer relatively cheaper products and more and more of such products should be recommended so that engagement is increased and also result in conversion rate which would finally result in increase in sales



-
- Additional data such as product category can improve the analysis so that it would give a better picture if people are inclined more towards a particular category when buying along with the price of the product

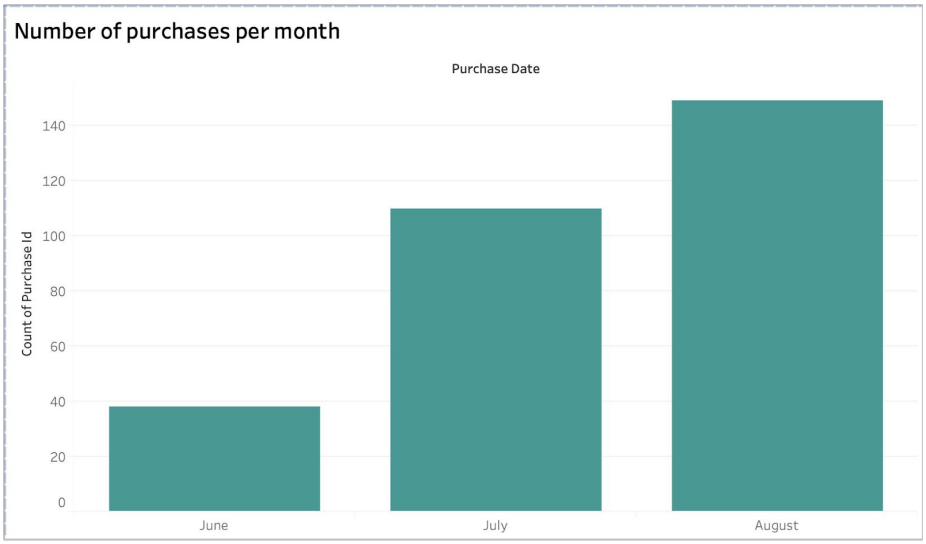
4. Number of Purchases per month growth

- Number of purchases is an important indicator as an increase in the number of purchases can result in an increase in sales and commission(not always!)

- As per the data, we can see that there is an increase in the number of purchases from June to July and from July to August. It is a good sign for the company. Definitely measures success.

- In our case, the number of purchases has increased but the sales amount has decreased from July to August. Hence, number of sales and sales amount are not always directly proportional. This can be due to factors such as users buying items which are relatively cheaper

-



5. Commission per month growth

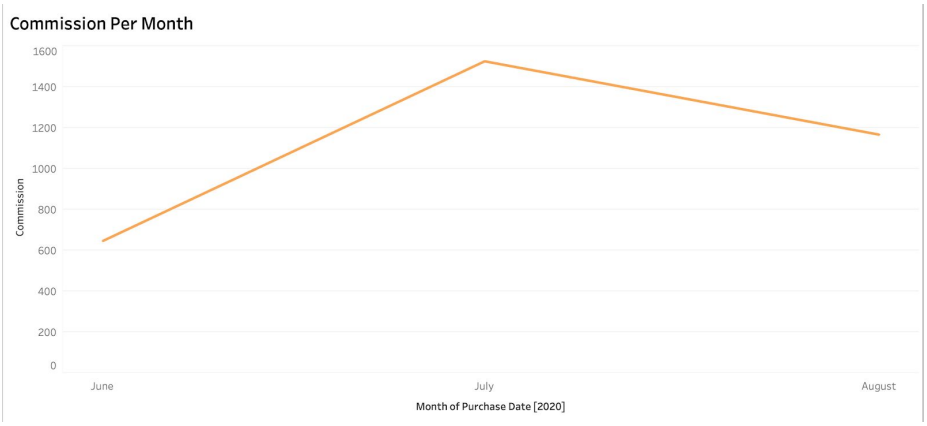
- One of the major sources of income for Qatch is through the Commission, hence finding the commission growth is important as you want to see how the company is performing with respect to its growth.

- The commission percentage for Qatch ranges from 5% to 20% as per the given data(found using : [commission*100/purchase amount])

- The total commission earned through June to July increased while it decreased from July to August. The number of purchases increased from July to August but the commission decreased. The reason for it can be due to users buying products whose commission percentage is low. Hence, decrease in commission should not be considered as a drop in performance as the number of purchases are increasing

- The company should aim at increasing the number of purchases. This is would in turn help in growth in commission in longer run

-



6. Reaction-Time of the day

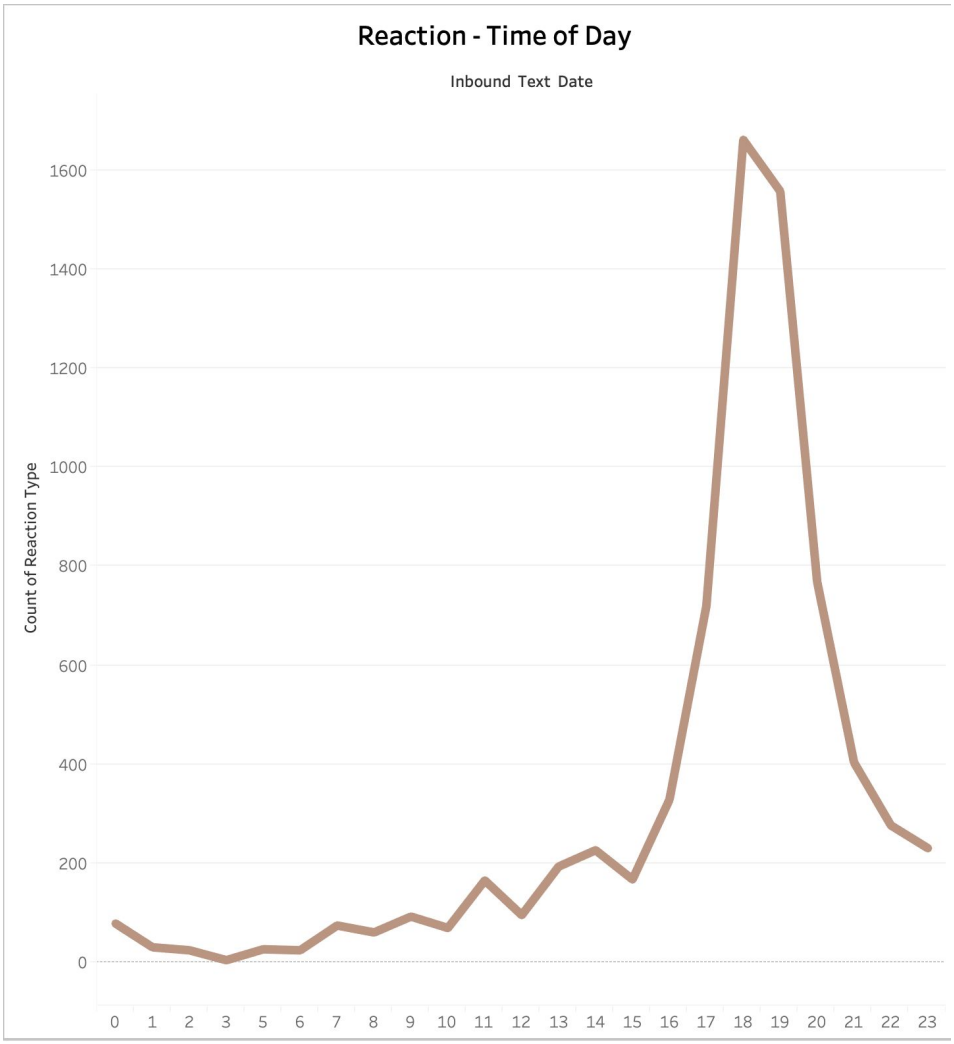
- Reaction-Time of the day is the plot of Number of Reactions to Time of the Day(Hours)

- Interestingly, the number of reactions increase from the 3pm hour mark, increase upto 6hr mark and then start dropping till the end of the day.

- Hence, it can be said that outbound_texts sent out during 3pm to 8pm can result in higher reaction rate hence improved engagement rate.

- This in turn might lead to an increased purchase rate with suitable recommendations.

- This is a KPI which acts as a suggestion and can contribute to the growth of the company if properly implemented. It gives us necessary insights on the when the messages should be sent to user to gain maximum reactions



-
- Providing information about the time at which outbound_text was sent can be useful. We can find correlation between time at which message was sent with time of reaction
- Also information about time of purchase can give useful insight of the time at which users are most likely to buy items. Hence recommendations can be sent out at the time where the user has the highest impulse of buying.

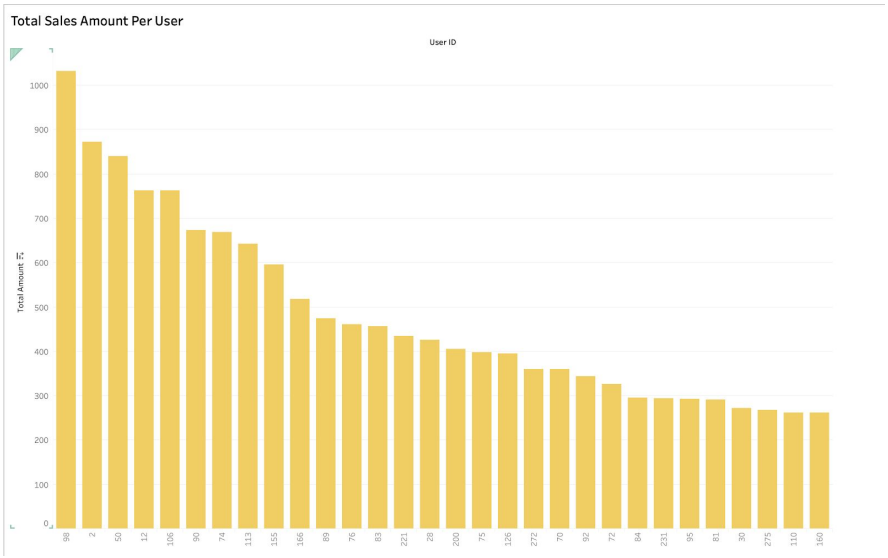
7. Total purchases and Total Sales per user

- Total purchases per user and Total Sales per user are two factors that give useful information to improve the company's success with respect to users.

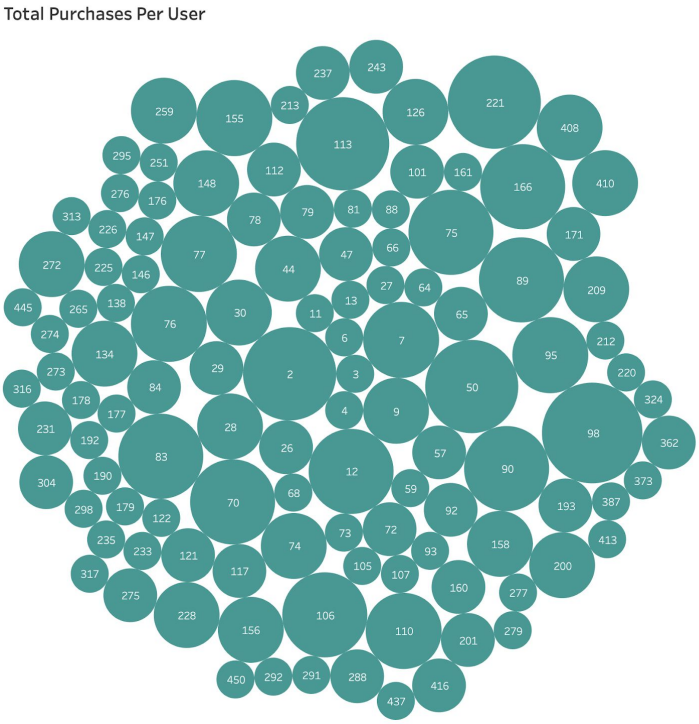
- Users with most sales and most purchases help in identifying the potential customers that are most likely to buy products. Hence providing promotional offers to these users can help in retention of these premium users as they help in the company’s growth.

- At the same time, identifying customers who are at the bottom tier in terms of sales and number of purchases should be segmented as low-value customers. These type of customers need a push from the company in terms of better recommendations or promotional offers
- Information about user demographics(age, place etc.) can help in identifying which kind of users are buying products and from where

-



-



-

- ReceivedToPurchaseRatio

```
select a.uid as UserID, count(a.text_id) as TotalReceived, count(b.purchase_id) as TotalPurchased
from
`qatch.outbound_text` a
left join
`qatch.purchases` b
on
a.text_id = b.outbound_text_id
group by UserID
order by UserID desc
```

- TotalAmountPerUser

```
select a.uid, sum(b.amount) as totalAmount
from `qatch.outbound_text` a
inner join
`qatch.purchases` b
on a.text_id = b.outbound_text_id
group by a.uid
order by totalAmount desc
```