**Business Analyst Assignment**

1. **To calculate the lifetime value (LTV) of the users acquired through different marketing channels**, we can use the following formula.

LTV = Average Revenue Per User (ARPU) \* Average user Lifespan

ARPU = Total Revenue / total number of users

Average user lifespan = Average(last login date – sign\_up date)

With sql, we can assume that we have created 3 tables as per the scheme naming:-

user\_sign\_up

user\_offer\_completion\_data

reward\_details

We can find the LTV by the following query

CREATE VIEW ltv\_table AS

SELECT

us.user\_id,

utm\_source,

us.created\_at,

last\_login\_at,

DATEDIFF(last\_login\_at, us.created\_at) AS lifespan,

SUM(rd.total\_revenue\_in\_paise) AS revenue

FROM user\_sign\_up us

JOIN user\_offer\_completion\_data uof

ON us.user\_id = uof.user\_id

JOIN reward\_details rd

ON uof.reward\_id = rd.reward\_id

GROUP BY user\_id;

SELECT

(SUM(revenue)/COUNT(DISTINCT user\_id)) \* AVG(lifespan) AS ltv from ltv\_table;

This query will allow us to calculate the lifetime value (LTV) of the users acquired through different marketing channels.

1. **To find insights from the data for both these apps and tell which app is better of these two.**

Steps

1. First we loaded the dataset from sql to get the combined data of the four tables.

We used the following query to perform the data collection.

SELECT

us.user\_id, us.app\_id, us.created\_at, us.last\_login\_at,

uod.offer\_id, uod.status as 'Offer\_status', uod.started\_at as 'offer\_started', uod.completed\_at as 'offer\_completed', uod.expires\_at as 'Offer\_expires',

uoc.reward\_id,uoc.created\_at as 'offer\_completed',

rd.label\_in\_english, rd.total\_payout\_in\_paise, rd.total\_revenue\_in\_paise

FROM users\_signup us

JOIN user\_offer\_data uod

ON us.user\_id = uod.user\_id

JOIN user\_offer\_completion\_data uoc

ON us.user\_id = uoc.user\_id

JOIN rewards\_details rd

ON uoc.reward\_id = rd.reward\_id;

Graphical user interface, text, application, email

Description automatically generated

Snippet of combined dataset from sql loaded to python

* The data currently includes 1000 records and 14 features.

1. Since we had many date features, we extracted the necessary information from each date feature and added to the dataset as new features.
2. We included few more features such as "Gross\_Profit", "Lifetime", "Reward\_completion\_time" to the dataset for further analysis.

Gross profit = Total revenue – Total payout

Lifetime = Last login – account created date

Reward completion time = Offer completed date – offer started date

This helped us for the easy analysis of data and provided more insights.

**Data Visualization for Analysis**

**Chart, bar chart, treemap chart

Description automatically generated**

**Chart, pie chart

Description automatically generated**

Number of records for each app

We can see that the number of records in data for each app users are different. This may cause bias effect. This can be rectified by balancing the data. One of the methods which can be used Is oversampling the records using SMOTE technique. But this can alter the true nature of the data. Thus we are proceeding with further analysis without balancing the data.

Chart

Description automatically generated with medium confidenceChart, line chart

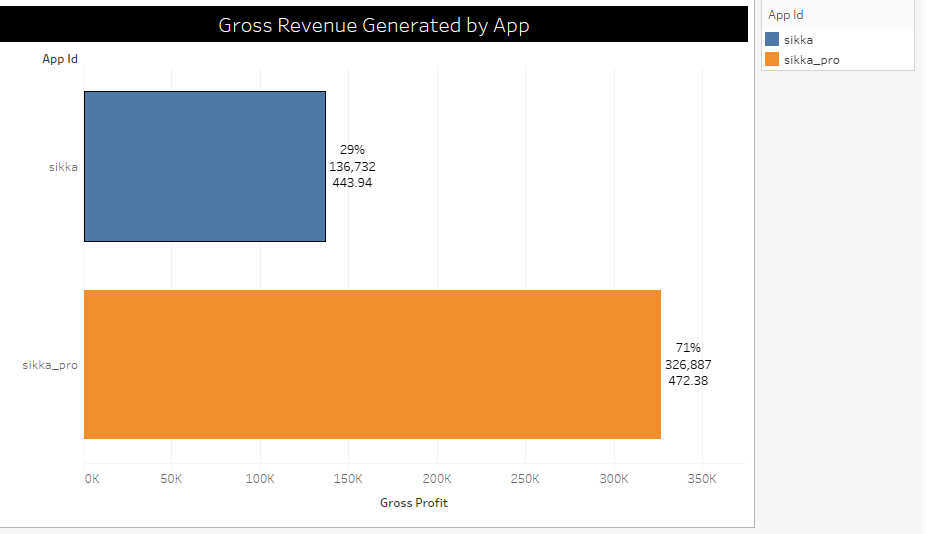
Description automatically generated

Total revenue and payout for each app

Time and days required for reward completion.

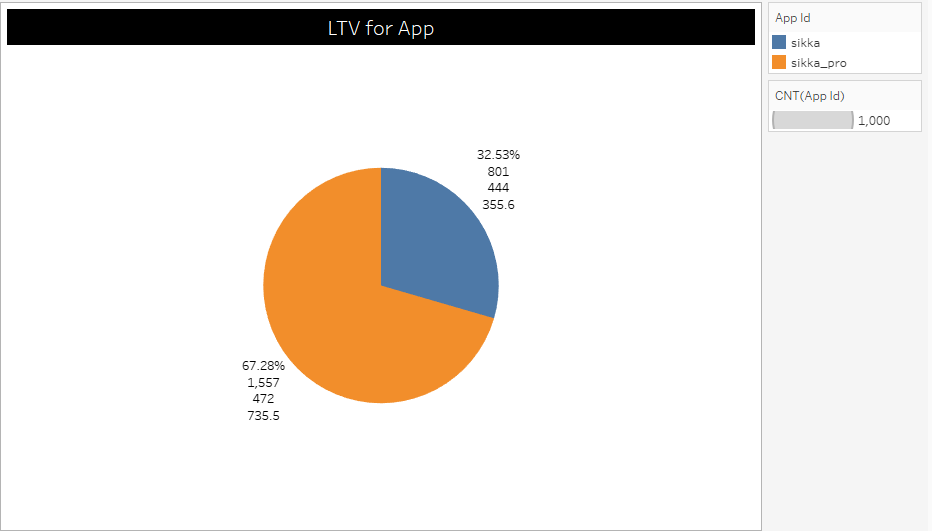
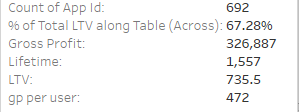
Observations

1. We can see that the data is provided more for sikka pro users(692) compared to sikka app users(308). So the data can be baised with the results.
2. The revenue and payout for both sikka and sikka pro is similar. Although there is an outlier for sikka app which can cause a bias in the data distribution. But majority of data points are lying in a normal distrubution.
3. The reward completion time is comparatively higher for sikka app compared to sikka pro. But no much difference are existing between the apps in reward completion time and days.



Gross Profit from users for each app

We can see the sikka pro app created more gross revenue (326,887) compared to the sikka app(136,732). But this can be biased since the sikka pro app users are more and it can show a higher aggregate revenue. So we calculated Gross profit per user for both apps and it clearly say the sikka pro app(472.38) indeed generate more revenue than the sikka app(443.94). It means the sikka pro is more beneficial for the app provider compared to sikka app.



Pie Chart of LTV for each app

We can see that the sikka pro app is covering 67.28% in total percent of LTV compared to the sikka app(32.53%).

Table

Description automatically generated

Summary of analysis

Insights:

1. The dataset includes 692 records for sikka pro user data and 308 records for sikka app user data.
2. The gross profit(revenue after user payout) is 326,887 for sikka pro app and 136,732 fir sikka app.
3. The gross profit per user for each apps clearly shows that the sikka pro app generates more revenue for the app provider from each users (472) than the sikka app(444)
4. While the revenue generation is more for sikka pro app, the reward completion time for each offer is more for sikka pro app(237) than the sikka app(164).
5. The sikka pro users are more likely to stay on the app than the sikka app users. The lifetime use of sikka app by the user(801) is less than the lifetime for sikka pro app(1557).
6. Finally when we compare the LTV it is clear that the Sikka pro app is generating more user life time value(735) than the sikka app(356). That is 67.28% ltv for sikka pro app and 32.53% ltv for sikka app.

**Conclusion**

After analyzing the following information.

Offer Initiation by users

Offer Completion by users

Rewards earned by users

Revenue generated

It is clear that the sikka pro app is generating more revenue and having more lifetime users. But the reward completion time for offers of sikka pro app is higher than the sikka app which can be reduced to increase more ltv for the app. **Overall the sikka pro app is better.**

1. **To predict the number of referrals for these 15 days of November**

Predicted results.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Date | DAU | Installs | Uninstalls | Signups | **Referrals** |
| 0 | 12816 | 3763 | 4149 | 1806 | **571.1113** |
| 1 | 12812 | 3087 | 3868 | 1550 | **522.1139** |
| 2 | 12042 | 3176 | 3815 | 1410 | **501.2439** |
| 3 | 12595 | 3172 | 3878 | 1629 | **523.5089** |
| 4 | 12361 | 3390 | 4021 | 1578 | **550.0537** |
| 5 | 13166 | 3441 | 4071 | 1656 | **563.4634** |
| 6 | 12565 | 3468 | 4011 | 1556 | **509.814** |
| 7 | 12988 | 4468 | 4143 | 1808 | **563.1845** |
| 8 | 12992 | 4491 | 4638 | 2017 | **586.3809** |
| 9 | 13377 | 4261 | 4480 | 1997 | **576.3156** |
| 10 | 13826 | 4274 | 4512 | 2047 | **582.6273** |
| 11 | 13464 | 4660 | 4856 | 2066 | **621.8167** |
| 12 | 13415 | 4416 | 4749 | 2147 | **617.0689** |
| 13 | 13873 | 4097 | 4305 | 2065 | **559.5537** |
| 14 | 14459 | 4890 | 4593 | 2707 | **645.481** |

1. **To find out if there is any anomaly present in the data for any of the apps present in the ADX sample dataset.**

**Graphical user interface

Description automatically generated**

App Ids with anomalies corresponding to the columns.

**Findings**

App ids with corresponding column of outlier data points.

* In the column ["requests", "ae\_response", "ae\_impressios"] the app ids ['0931528', '43840611', '47862278','63338696', '66981243'] are having extreme outlier data points.
* In the column 'ae\_clicks' the app id "66981243" is having extreme outlier data points.
* In the column, "ae\_revenue" the app ids ['16243270', '40931528'] are having extreme outlier data points.
* In the column, "show-rate". almost all app ids are having similar data point distribution some outlier data points.
* In the column, "click-rate", the app ids ['16243270', '95674771', '25596802', '63338696'] are having extreme outlier data points.