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.

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

Steps

a. 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;
```

	user_id	app_id	created_at	last_login_at	offer_id	Offer_status	offer_started	offer_ended	Offer_expires	reward_id	offer_c
0	44f9c9f1-1ce1-4b04-8042-ddb1c548c589	sikka	2022-11-26	2022-11-29	3529	ONGOING	2022-11-29 01:14:50.965031+00:00	NaT	2022-11-30 18:29:59+00:00	935	2
1	44f9c9f1-1ce1-4b04-8042-ddb1c548c589	sikka	2022-11-26	2022-11-29	3479	ONGOING	2022-11-29 07:46:21.644530+00:00	NaT	2022-12-01 07:46:21.677847+00:00	935	2
2	44f9c9f1-1ce1-4b04-8042-ddb1c548c589	sikka	2022-11-26	2022-11-29	3549	ONGOING	2022-11-29 08:01:03.045611+00:00	NaT	2022-12-01 08:01:03.050394+00:00	935	2
3	44f9c9f1-1ce1-4b04-8042-ddb1c548c589	sikka	2022-11-26	2022-11-29	3443	COMPLETED	2022-11-28 08:02:55.130103+00:00	2022-11-28 08:06:39.074543+00:00	NaT	935	2
4	44f9c9f1-1ce1-4b04-8042-ddb1c548c589	sikka	2022-11-26	2022-11-29	3431	ONGOING	2022-11-28 08:01:56.418620+00:00	NaT	2022-11-30 08:01:56.422570+00:00	935	2

```
data.shape
```

(1000, 14)

Snippet of combined dataset from sql loaded to python

- The data currently includes 1000 records and 14 features.

b. Since we had many date features, we extracted the necessary information from each date feature and added to the dataset as new features.

c. We included few more features such as "Gross_Profit", "Lifetime", "Reward_completion_time" to the dataset for further analysis.

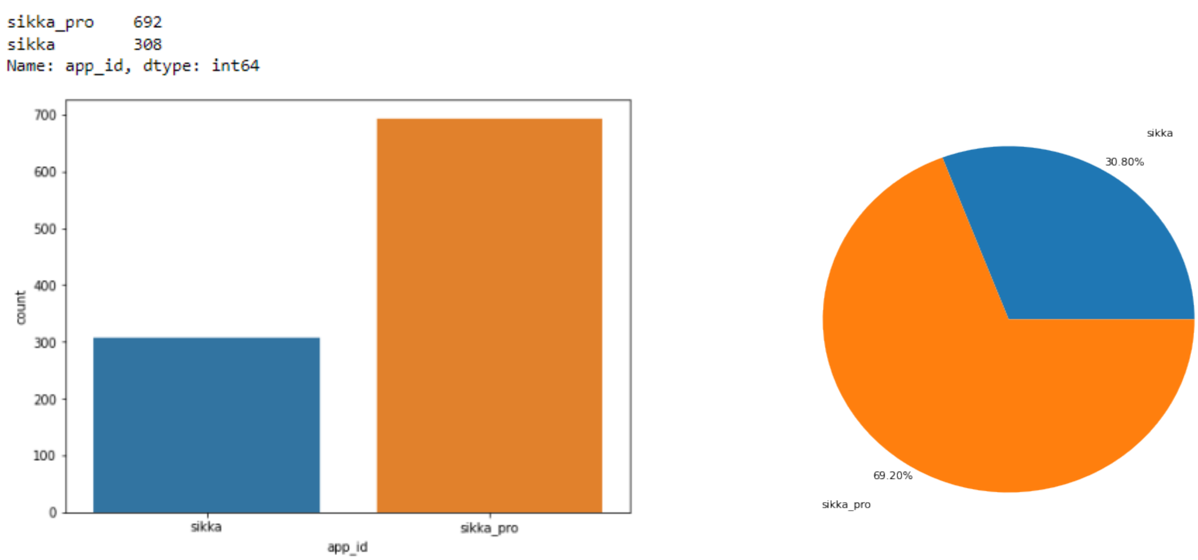
Gross profit = Total revenue – Total payout

Lifetime = Last login – account created date

Reward completion time = Offer completed date – offer started date

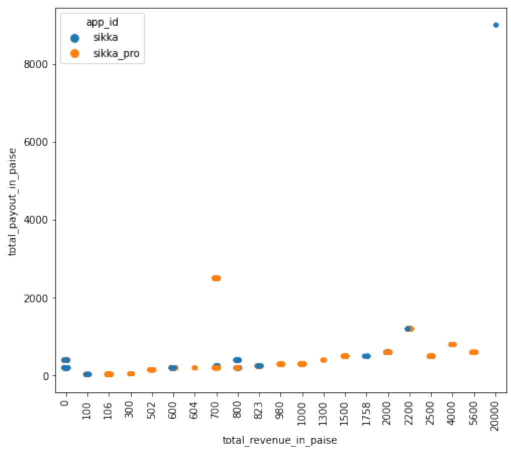
This helped us for the easy analysis of data and provided more insights.

Data Visualization for Analysis

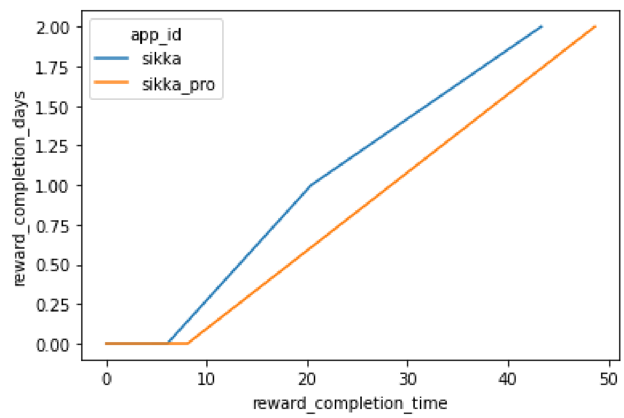


Number of records for each app

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.



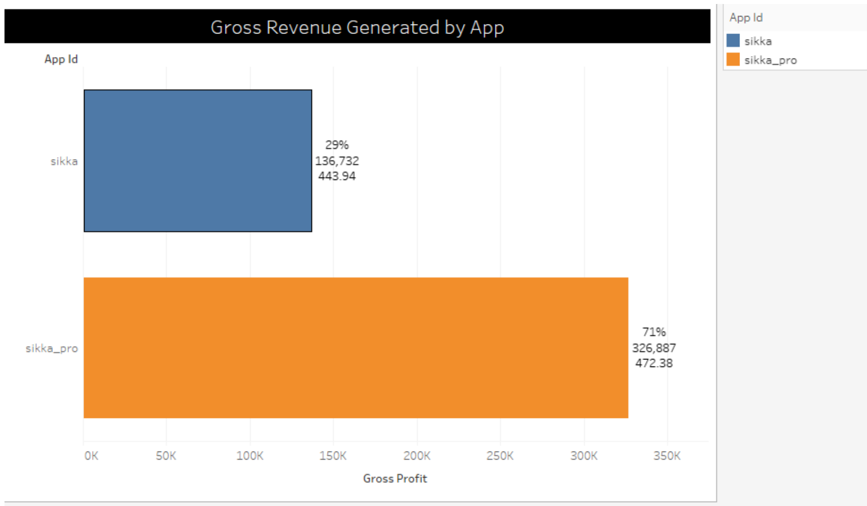
Total revenue and payout for each app



Time and days required for reward completion.

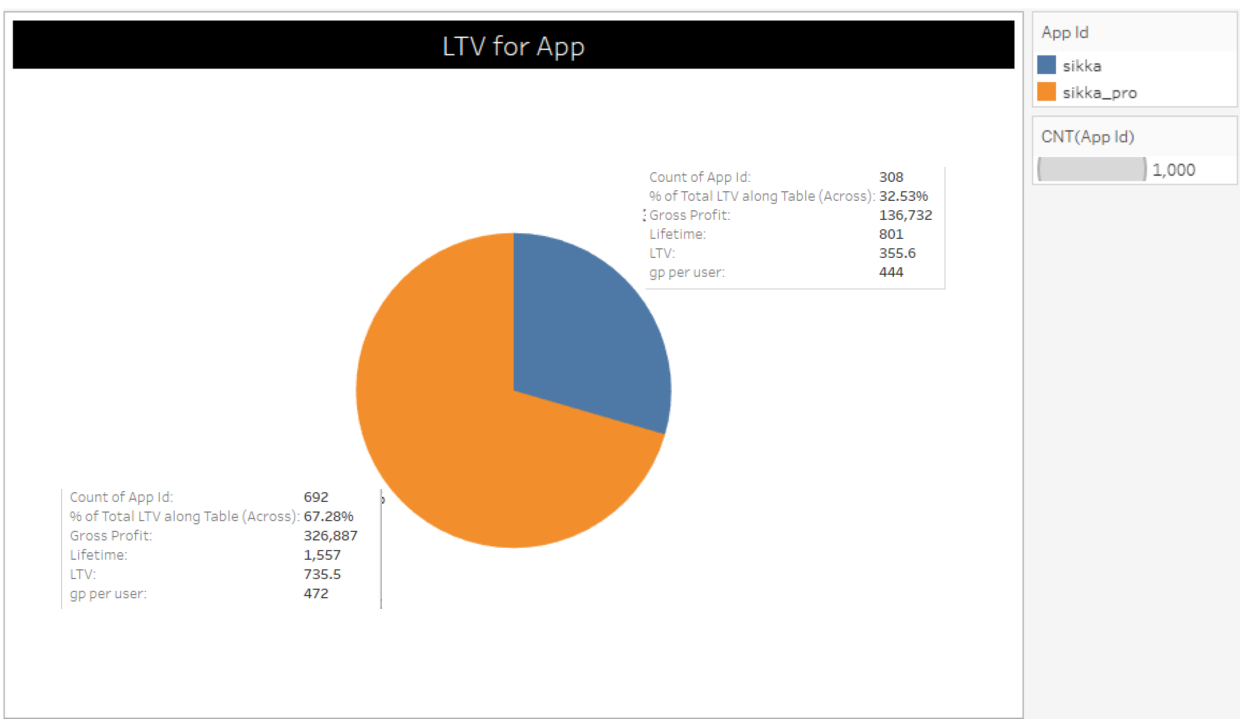
Observations

a. 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.
b. 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.
c. 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%).

Summary of Analysis							
App Id	Count of App Id	Gross Profit	gross profit to revenue ratio	Reward Completion Time	Lifetime	LTV	% of Total LTV along Table (Down)
sikka	308	136,732	13,511	164	801	356	32.53%
sikka_pro	692	326,887	28,714	237	1,557	735	67.28%
Grand Total	1,000	463,619	42,225	400	2,358	1,093	100.00%

Summary of analysis

Insights:

a. The dataset includes 692 records for sikka pro user data and 308 records for sikka app user data.
b. The gross profit(revenue after user payout) is 326,887 for sikka pro app and 136,732 fir sikka app.
c. 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)
d. 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).
e. 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).
f. 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

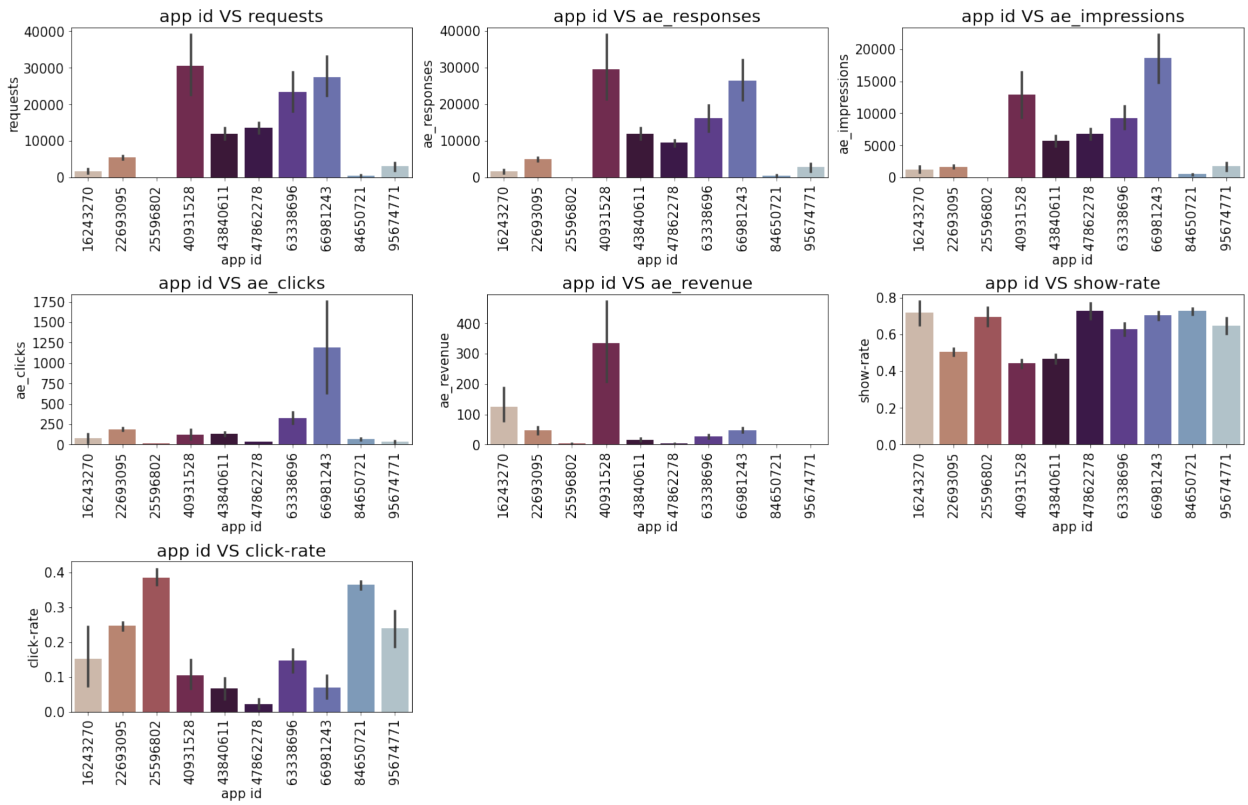
It is clear that the sikka pro app is generating more revenue and having more lifetime users. But the reward completion time for offers of sikka pro app is higher than the sikka app which can be reduced to increase more ltv for the app. **Overall the sikka pro app is better.**

3. 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**

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



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.