

REPORT

By


Vishwajeet Oraon

01/09/2025

[✉Email](#) [Website](#)



INDEX

SR NO.	CONTENT	PAGE NO.
1.	ASSIGNMENT DELIVERABLES AND PREREQUISITES	2
2.	PLAYER FUNNEL AND CONVERSION RATES	2-7
3.	RETENTION AND ENGAGEMENT	8-12
4.	PLAYER SEGMENTATION	13-18
5.	APPENDIX FOR SQL FILES, DASHBOARDS  Lognormal Business Analyst Assi...	19

ASSIGNMENT DELIVERABLES

Funnel & Conversion

1. Build the player funnel: Registrations → First Deposit → First Bet → Active in first 30 days.
 - Calculate conversion % at each stage.
 - Identify which stage has the largest drop-off and where (e.g., acquisition channel, cohorts).

Retention & Engagement

1. For each player, count days active in the first 30 days after registration.
 - Group players into cohorts (e.g., 1-2, 3-5, 6-10, etc.).
 - Which cohort contributes the most to total deposits, and what could explain that distribution?
2. Measure the time gap (days) between first deposit and first bet.
 - Report the mean, median, 75th percentile, and maximum values.
 - How does the distribution shape (short vs. long gaps) affect engagement outcomes?

Player Segmentation

1. Identify the top 10% of players by total deposit amount.
 - Compute their share of total deposits (%).
 - What does this tell you about the concentration of deposits in the player base?
2. Create a visual representation of first deposit amounts by binning them into meaningful buckets.
 - Do you observe clustering around certain values? Why might that occur in datasets generally?
 - Can first deposit amount indicate anything about a customer's profitability?

DATASETS

- Player details
- First bet data
- First deposit data
- Player activity
- Bonus cost data

TOOLS AND TECHNOLOGIES USED

1. PostgreSQL : for data cleaning and Exploratory Data Analysis
2. Power BI : for advanced data visualization
3. Python + Jupyter Notebook : for data manipulation

METHODOLOGY

1. Importing Data into PostgreSQL, validating the data
2. Exploratory data analysis to get descriptive statistics
3. Importing data into Power BI, building data model, creating visuals

1. FUNNEL AND CONVERSION

OBJECTIVE: TO TRACK PLAYER THROUGH FUNNEL AND IDENTIFY DROP OFFS

FUNNEL STAGES:

- STAGE 1: SIGNUP
- STAGE 2 : DEPOSIT
- STAGE 3 : FIRST BET
- STAGE 4 : ACTIVE WITHIN 30 DAYS OF SIGNUP

STEPS:

- Joined the tables using the src_player_id column
- Treated rows with NULL values in deposit/bet columns as no deposit/bet
- Counted players 'active within 30 days of signup' if first bet <= 30 days after signup
- **Assumption:** bets with amount = 0 (promotions/free bets) were not included as "valid bets" for funnel progression.

CONVERSION AT EACH STAGE:

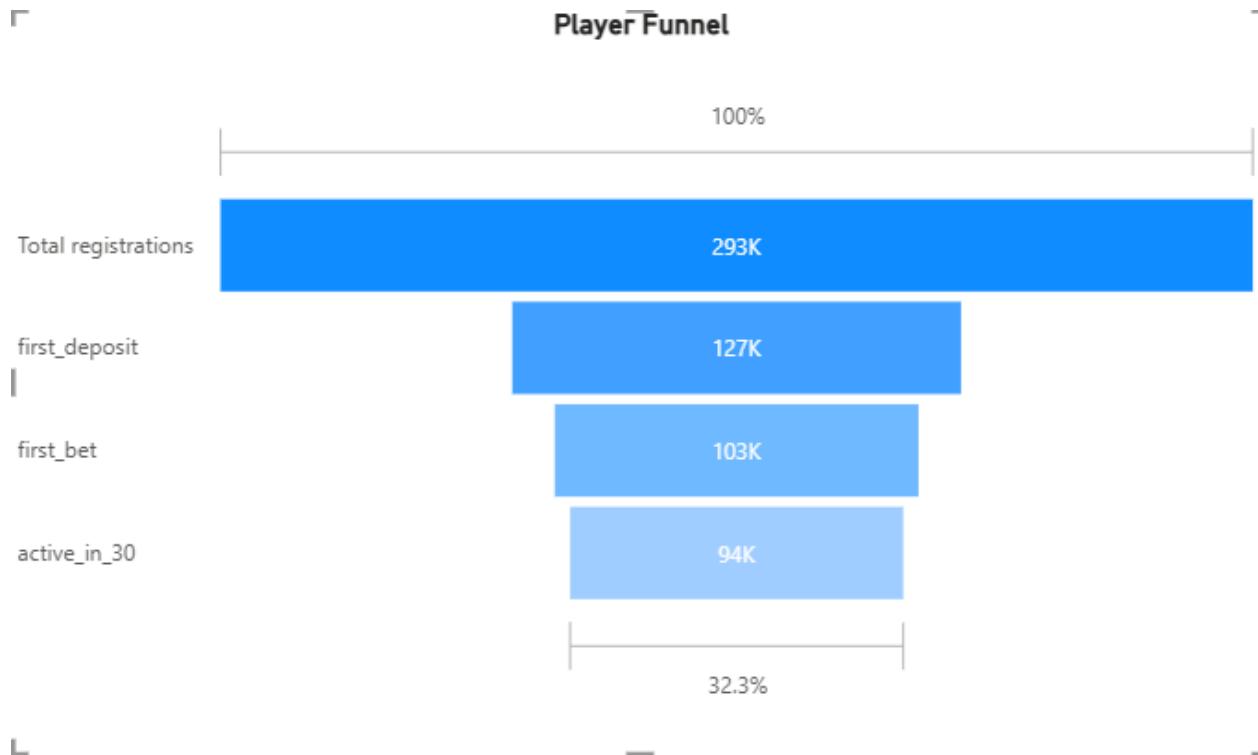
SQL QUERY:

```
Query Query History
1
2
3 WITH registrations AS (
4     SELECT src_player_id, signup_date
5     FROM player_details
6 ),
7 first_deposits AS (
8     SELECT src_player_id, first_deposit_date
9     FROM first_deposit
10    WHERE first_deposit_date IS NOT NULL
11 ),
12 first_bets AS (
13     SELECT src_player_id, system_first_bet_datetime, system_first_betslip_amt
14     FROM first_bet
15    WHERE system_first_bet_datetime IS NOT NULL AND system_first_betslip_amt > 0
16 )
17 SELECT
18     COUNT(DISTINCT r.src_player_id) AS registrations,
19     COUNT(DISTINCT fd.src_player_id) AS first_deposit,
20     COUNT(DISTINCT fb.src_player_id) AS first_bet,
21     COUNT(DISTINCT reg.src_player_id) AS active_in_30,
22     ROUND(100.0 * COUNT(DISTINCT fd.src_player_id) / COUNT(DISTINCT r.src_player_id), 2) AS reg_to_deposit_rate,
23     ROUND(100.0 * COUNT(DISTINCT fb.src_player_id) / NULLIF(COUNT(DISTINCT fd.src_player_id),0), 2) AS deposit_to_bet_rate,
24     ROUND(100.0 * COUNT(DISTINCT reg.src_player_id) / NULLIF(COUNT(DISTINCT fb.src_player_id),0), 2) AS first_bet_in_30
25 FROM registrations r
26 LEFT JOIN first_deposits fd ON r.src_player_id = fd.src_player_id
27 LEFT JOIN first_bets fb ON fd.src_player_id = fb.src_player_id
28 LEFT JOIN registrations reg ON fb.src_player_id = reg.src_player_id AND fb.system_first_bet_datetime <= reg.signup_date + INTERVAL '30 days' ;
29
```

	registrations bigint	first_deposit bigint	first_bet bigint	active_in_30 bigint	reg_to_deposit_rate numeric	deposit_to_bet_rate numeric	first_bet_in_30 numeric
1	292785	127261	103146	94457	43.47	81.05	91.58

*The sql file is attached with the report.

FINDINGS:



- Registration to deposit : **43%** , Less than half of registered users deposit, which is a critical drop-off point.
- Deposit to bet : **81%** , Strong conversion, most of them go on to place a bet
- First bet within 30 days of activity : **91%**, most of them place their first bet within 30 days of signup

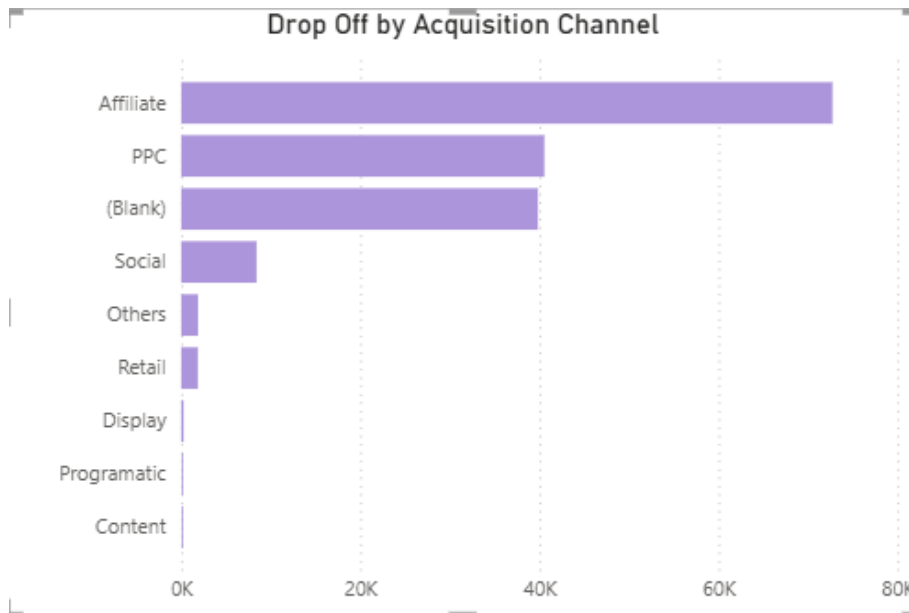
KEY INSIGHTS ON DROP OFF

- **Biggest Drop-off:** Registration to Deposit (**only ~43%**). This indicates a need to improve onboarding, deposit incentives, or trust-building measures.

- **DROP OFF BY ACQUISITION CHANNEL**

	acquisition_channel character varying (20) 🔒	count bigint 🔒
1	Affiliate	72731
2	PPC	40554
3	[null]	39835
4	Social	8410
5	Others	1874
6	Retail	1853
7	Display	260
8	Programatic	6
9	Content	1

*SQL file is attached with the report.

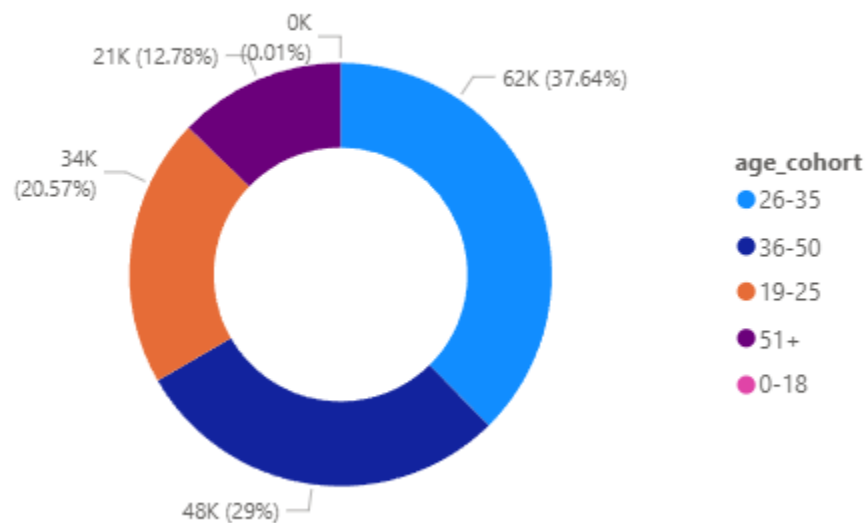


- **Insight:** **Affiliate + PPC** combined form **~80% of total drop-offs**

- **DROP OFF BY AGE**

	age_cohort text	count bigint
1	36+	69158
2	26-35	62304
3	18-25	34050
4	<18	12

Drop off by age by age_cohort



Observations:

- The **largest number of drop-offs** comes from the **36+ cohort (69,158)**, followed by 26–35.
- The **18–25 cohort** also contributes significantly (34,050).
- <18 drop-offs are negligible.

Insight: The drop-off is heavily skewed toward older users simply because they are higher in numbers too

2. RETENTION AND ENGAGEMENT

OBJECTIVE:

- Measure player activity and engagement within the first 30 days after registration.
- Identify which activity cohorts contribute most to total deposits.
- Analyze the time gap between first deposit and first bet to understand engagement patterns.
- Draw insights on how engagement duration and deposit-to-bet timing affect overall player activity and retention.

DATA LIMITATION

Player activity is available as monthly aggregates, so exact days active in the first 30 days post-registration cannot be determined as a 30 day window will fall on two months.

APPROACH

- Since the activity data is monthly, I estimated the first 30-day active days by interpolating proportionally. For example, if a player signed up on 6th Jan had 20 active days in Jan then this means he was active on 20 days out of 24 days. Hence, their 30 day estimate is given by $20 \times 30 / 24$.
- After the approximation, created active days cohort like 0-7(first week), 8-15(second week) and so on.

SQL QUERY:

Query
Query History

```

6
7 WITH activity AS (
8     SELECT
9         pd.src_player_id,
10        pd.signup_date,
11        pa.activitymonth,
12        pa.activeplayerdays,
13        EXTRACT(DAY FROM pa.activitymonth) - EXTRACT(DAY FROM pd.signup_date) + 1 AS rem_days_in_month
14    FROM player_details pd
15    JOIN player_activity pa
16    ON pd.src_player_id = pa.src_player_id
17    AND DATE_TRUNC('month', pd.signup_date) = DATE_TRUNC('month', pa.activitymonth)
18 )
19 SELECT
20     src_player_id,
21     FLOOR((activeplayerdays :: numeric/rem_days_in_month)*30) AS estimated_activity
22 FROM activity;
23

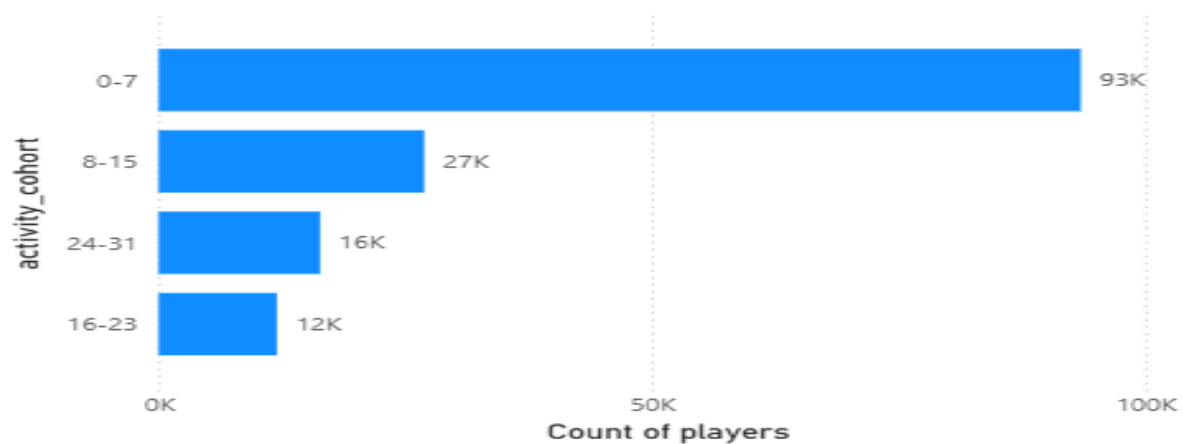
```

Data Output
Messages
Notifications

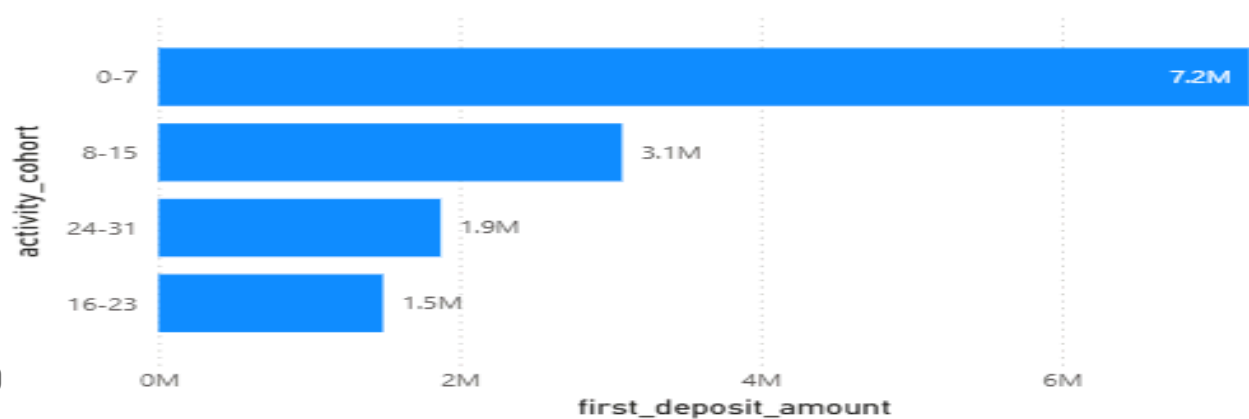
Showing rows: 1 to 1000
Page No: 1

	src_player_id integer	estimated_activity numeric
1	13519049	26
2	13395231	22
3	18017452	3
4	15095299	2
5	14775145	1
6	14827736	3

Count of players by activity_cohort



first_deposit_amount by activity_cohort



activity_cohort	count	total_deposit	depoit_to_count_ratio
0-7	57752	7248860	125.52
16-23	11926	1490644	124.99
24-31	15159	1874503	123.66
8-15	24172	3079331	127.39

FINDINGS:

- 93K players where active within 30 days of signup
- The 0-7 cohort has the most players which makes 7.2M deposit
- Also investigated whether there's any disproportionality, but the deposit to count ratio for the all cohorts were nearly same, indicating every cohort has deposited in proportion to their count.

Measuring the time gap between first deposit and first bet

- Calculated the gap between the first deposit date and first bet using the first_deposit date and system_first_bet_datetime column
- Created a cohort of gaps like 0(on the same day), 0-7(in a week), (7-30) within a month
- **There were 3710 rows** where the first bet was placed before the first deposit, indicating some kind joining bonus or free bet, so those rows were not taken into account

SQL QUERY:

Query

Query History

1

2

3

4

5

6

7

8

9

10

11

12

WITH CTE AS (
SELECT src_player_id, first_deposit_date
FROM first_deposit fd
WHERE first_deposit_date IS NOT NULL
)
SELECT CTE.src_player_id, first_deposit_date, system_first_bet_datetime,
DATE(system_first_bet_datetime) - first_deposit_date AS Gap
FROM CTE
LEFT JOIN first_bet fb
ON CTE.src_player_id = fb.src_player_id
WHERE system_first_bet_datetime IS NOT NULL AND system_first_betslip_amt > 0
AND DATE(system_first_bet_datetime) - first_deposit_date >= 0;

Data Output

Messages

Notifications

SQL

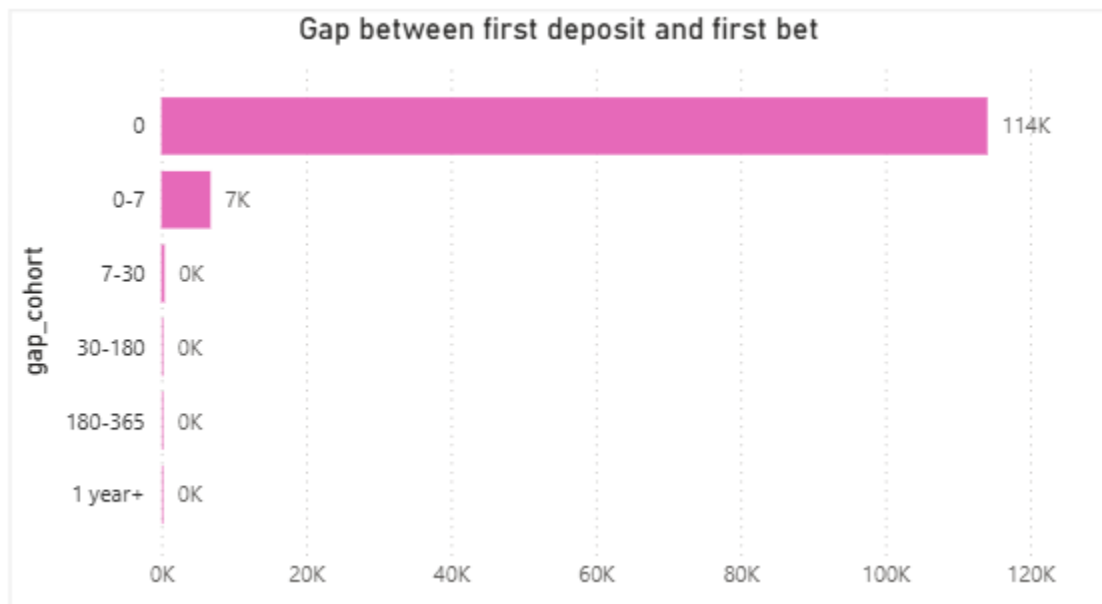
Showing rows: 1 to 1000

Page No: 1

	src_player_id integer	first_deposit_date date	system_first_bet_datetime timestamp without time zone	gap integer
1	20424535	2021-06-05	2021-06-05 16:12:00	0
2	20538009	2021-06-16	2021-06-16 21:07:00	0
3	20528494	2021-06-15	2021-06-15 23:46:00	0
4	14918789	2020-02-07	2020-02-07 19:35:00	0
5	12547287	2020-12-12	2020-12-12 21:06:00	0
6	17827614	2020-11-17	2020-11-17 17:10:00	0
7	12791782	2019-04-08	2019-04-08 18:24:00	0
8	15435540	2020-09-20	2020-09-20 22:56:00	0
9	13188473	2019-06-23	2019-06-23 14:42:00	0

FINDINGS

- The data of gap of was highly skewed, **90% of the players** placed their bet on the date of deposit itself



- Because of the highly skewed nature of the gaps, the mean, median and the 75th percentile were around 0, where as the maximum gap was of 728 days.

	mean numeric	median integer	percentile_75 integer	max integer
1	0.28	0	0	728

SHORT VS LONG GAPS: AFFECT ON ENGAGEMENTS OUTCOMES

- Short gap shows player has high intent. They had the intention of placing bet right after depositing. These players are more likely to be active and have high life time value
- Long gaps player is hesitant or cautious. Higher chance of drop off before first bet. Even if they bet, frequency may be lower.

3. PLAYER SEGMENTATION

OBJECTIVE:

- To identify top 10% of players by total deposit
- To calculate their share in the total deposit
- To gain insight on the concentration of deposits in the players base
- To create a visual representation of first deposit amounts by creating suitable buckets like (0-100,100-500,500-1000,1000-1000, etc)
- Observe the clustering around certain values and the find the possible reasons
- To investigate if the players first deposit amount indicate anything about customer's profitability

APPROACH:

- Used the first deposit table to find out top 10 % of the players by ordering them by their first deposit amount
- To gain information about the concentration, investigated against pareto principle
- Created suitable bins to see the amount of money players generally make as first deposit
- To gain insight about the profitability of player based on their first deposit, further looked into their activity

1. Top 10% of players by total deposit

SQL QUERY:

QueryQuery History

```
21
22 -- Top 10% of players by total deposit amount
23 SELECT *
24 FROM First_deposit
25 WHERE first_deposit_amount IS NOT NULL
26 ORDER BY first_deposit_amount DESC
27 LIMIT (SELECT FLOOR(0.1*COUNT(*)) FROM First_deposit
28 WHERE first_deposit_amount IS NOT NULL);
29
30
```

Data OutputMessagesNotifications

SQL

Showing rows: 1 to 1000

	src_player_id integer	first_deposit_date date	first_deposit_channel character varying (20)	first_deposit_method character varying (20)	first_deposit_amount integer
1	12331133	2021-05-15	Retail	RetailDeposit	13000
2	13070582	2019-05-23	Online	VISA	10000
3	13906563	2019-11-19	Retail	RetailDeposit	10000
4	12472509	2020-02-16	Retail	RetailDeposit	10000
5	18021026	2020-11-25	Online	MC	9184
6	13660399	2019-10-15	Online	MoneyBookersViaSC	7000
7	19823834	2021-04-09	Retail	RetailDeposit	7000
8	13654834	2019-10-11	Online	MoneyBookersViaSC	7000
9	17623038	2020-11-01	Retail	RetailDeposit	6615
10	19096154	2021-02-12	Online	VISA	6000
11	18337436	2020-12-12	Retail	RetailDeposit	5500
12	12572931	2019-02-21	Online	VISA	5000
13	17050710	2020-11-20	Online	MC	5000

Total rows: 12726Query complete 00:00:04.610

SOME STATISTICS OF TOP 10% PLAYERS BY DEPOSITS

Total deposit of Top 10

7.58M

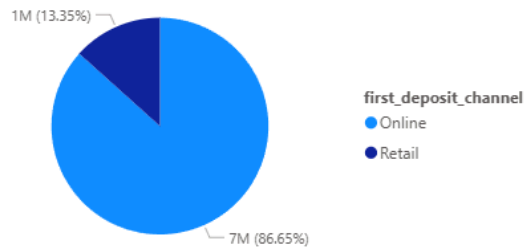
Total deposits

16M

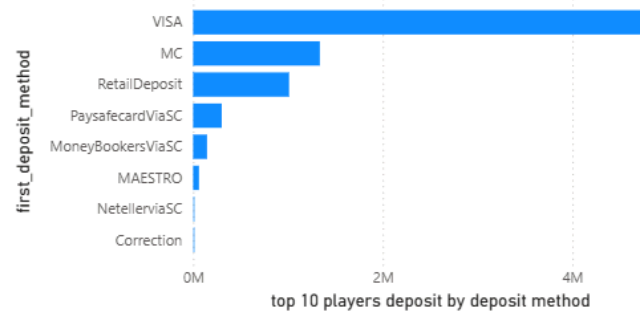
deposit_share

48.58

top 10 players deposit by deposit channel by first_deposit_channel



top 10 players deposit by deposit method by first_deposit_method



FINDINGS:

- Top 10% players i.e 12726 out of 127261 contributed to **49% of the total deposits**
- 86% of them deposited through the online channel
- Most of them used the VISA as their deposit method

INVESTIGATED AGAINST THE PARETO PRINCIPLE

SOME STATISTICS OF TOP 20% PLAYERS BY DEPOSITS

Total deposit of top 20%

10M

Total deposit of Top 10

7.58M

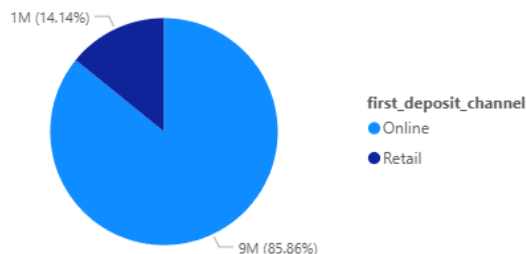
Total deposits

16M

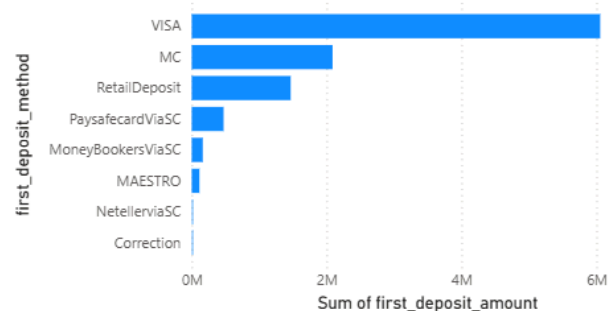
Deposit share of top 20%

66.48

Sum of first_deposit_amount by first_deposit_channel



Sum of first_deposit_amount by first_deposit_method



2. First deposit amount cohort

SQL QUERY:

Query

Query History

```
31
32 WITH CTE AS (
33     SELECT *
34     FROM First_deposit
35     WHERE first_deposit_amount IS NOT NULL)
36 SELECT src_player_id, First_deposit_amount,
37 CASE
38     WHEN First_deposit_amount BETWEEN 0 AND 100 THEN '0-100'
39     WHEN First_deposit_amount BETWEEN 101 AND 500 THEN '101-500'
40     WHEN First_deposit_amount BETWEEN 501 AND 1000 THEN '501-1000'
41     WHEN First_deposit_amount BETWEEN 1001 AND 5000 THEN '1001-5000'
42     WHEN First_deposit_amount BETWEEN 50001 AND 10000 THEN '5001-10000'
43     WHEN First_deposit_amount BETWEEN 10001 AND 20000 THEN '10001-20000'
44     END AS deposit_bins
45 FROM CTE;
46
47
```

Data Output

Messages

Notifications

+

📄

▼

📋

▼

🗑️

📦

📥

📶

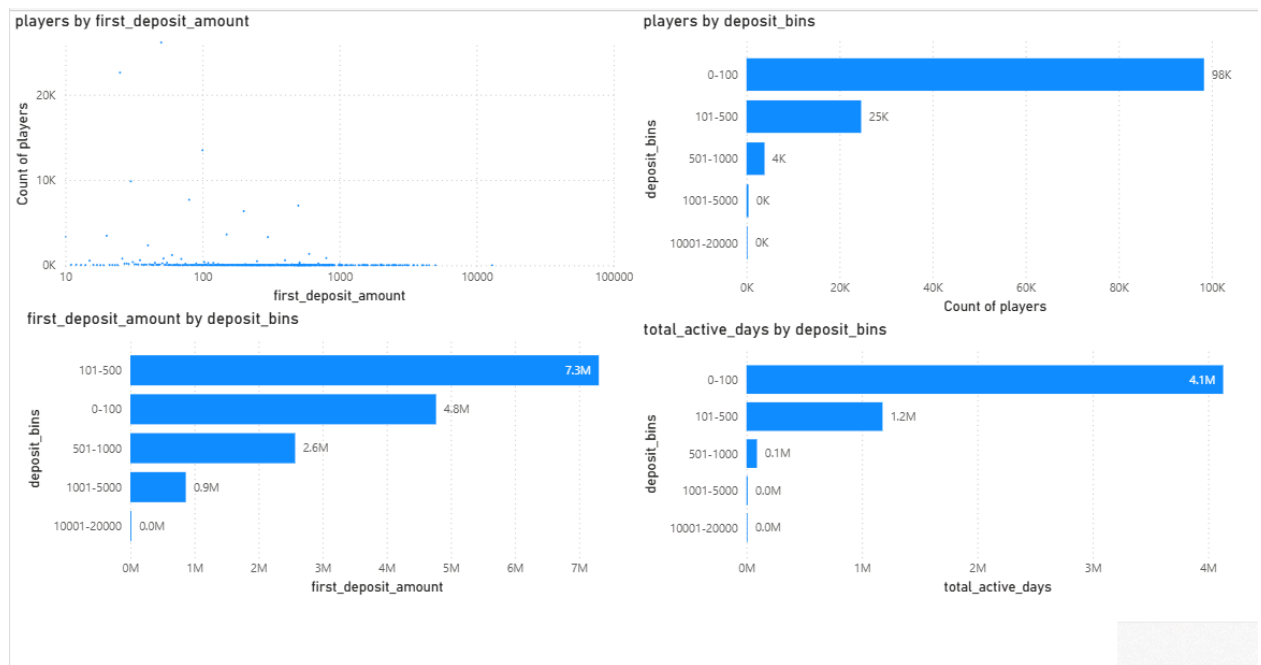
SQL

Showing rows: 1 to 1000

	src_player_id integer	first_deposit_amount integer	deposit_bins text
1	20424535	50	0-100
2	20538009	50	0-100
3	16605812	25	0-100
4	20528494	25	0-100
5	14918789	25	0-100
6	12547287	52	0-100
7	17827614	25	0-100

Total rows: 127261 Query complete 00:00:01.415

Charts giving insights based on the first deposit bins:



FINDINGS:

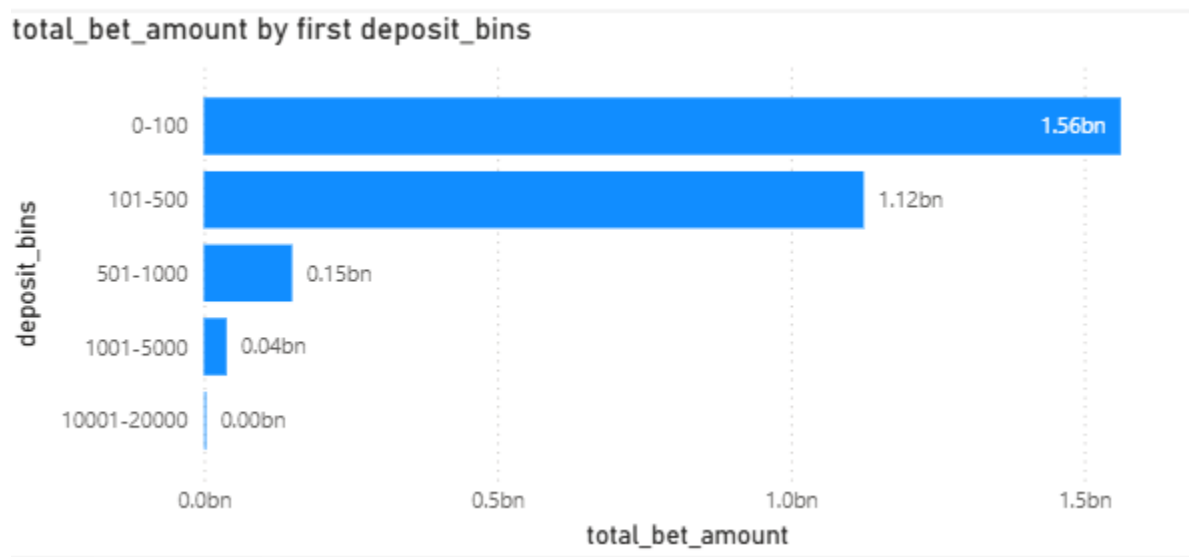
- 0-100 bin dominates, 98k players, following with 101-500 with 25k players
- After that a sharp drop off, only 4k in 501-1000 and almost negligible in 1000+
- **Highest no of players in 0-100 bin but 101-500 with fewer players has higher monetary value**
- **Engagement is driven by the low deposit mass (0-100) but the revenue is driven by the mid deposit segment (101-500)**

WHY CLUSTERING?

- User risk appetite: Most first time users start with a small safe deposit
- Platforms deposit thresholds: platforms often have set amounts (100,500,etc) so players cluster around those

WHAT DOES THE FIRST DEPOSIT AMOUNT TELL ABOUT A CUSTOMER PROFITABILITY?

To answer this, I went to check the players activity against each first deposit, like for the players who started with their deposit in 0-100, the total amount of bets they have placed.



FINDINGS:

- 0-100 depositors contribute 1.56bn total bet amount. This suggest that even low first depositors can remain active and generate high betting volume in the long run due to their numbers
- 101-500 depositors have second highest betting volume
- First deposit alone is not a perfect predictor of profitability, profitability seems driven by volume of players more than deposit size.

APPENDIX

Google drive link containing SQL files: [Lognormal Business Analyst Assignment](#)