**SUBMISSION INSTRUCTIONS**

Please submit through the greenhouse link.

**SQL Exercise**

For a planned CRM campaign, FanDuel Sportsbook is going to send an email to existing players letting them know the available markets and new site features for the upcoming NFL season. In addition there will be targeted offers to certain players based on their transactional history.

As an analyst your job is to help define the segmentation and then use the SQL database to generate lists of users. The data tables you have at your disposal are below:

**USER TABLE** – This is where we store information about user’s account(one row per customer)

| FIELD | DESCRIPTION |
| --- | --- |
| playerid | Unique identifier for customer |
| alias | Username of customer on site |
| email | Email address of customer |
| registration\_date | Date of customer’s registration |
| activation\_date | Date of customer’s first activity with their own money (i.e. after a deposit) |
| account\_balance | Snapshot of latest customer balance |

**ACTIVITY TABLE** – This is where we store information about user’s activity

| FIELD | DESCRIPTION |
| --- | --- |
| playerid | Unique identifier for customer |
| state | “NJ” or “PA” |
| betplaceddate | Date of bet placement |
| betsettleddate | Date of bet settlement |
| betid | Unique identifier of bet |
| legid | Unique identifier of leg (might be more than one per bet) |
| bettype | Bet type: can be Straight (1 leg) or Parlay (2+ legs) |
| stake | Amount wagered per leg |
| winning | Amount won per leg |
| ggr | Gross Revenue. Formula: Stake – Winning (+ if customer lost, - if customer won) |
| sportname | Sport selected per leg (Football, Basketball, etc.) |

**DEPOSIT TABLE** – This is where all the information about every deposit

| FIELD | DESCRIPTION |
| --- | --- |
| playerid | Unique identifier for customer |
| cartid | Unique identifier for deposit |
| payment\_status | Status of payment: “a” if approved, “d” if declined |
| payment\_amount | Amount deposited |
| cartdateid | Date of deposit |
| sb\_flag | Flag to identify Sportsbook deposits: 1 if SB, 0 if non-SB |

1. Based upon the information you have in the database, what variables would you consider most when designing a segmented offer for users in the new football season and why do you think that they are important? What sort of groups would you like to create and how might you vary the incentives/offer to each of those groups?

When preparing offers for a new football season I would probably start off with a general RFM analysis (Recency, Frequency, and Monetary Value). That will provide a baseline for the engagement of each customer and the value that they provide. Also it will segment the consumer base into groups of players who likely share the same motivations on how, when, and why they bet, making targeted campaigns more personalized and likely to result in a positive return on investment.

The fields that would be necessary to perform the RFM would be: R = Max(betplaceddate), F =COUNT(DISTINCT betid), and M =SUM(stake) . The three of would all be paired with playerid and I would do so by making a view of each the R, F, and M than bring them all together into a final view to create the RFM scoring

\* When doing the SUM(stake you would have to ensure that you are only getting one per distinct betid to ensure that duplicates from parlays are not over represented.\*

1. The CRM Manager has an idea for a test, but wants to know if there are enough users available to make the results significant, so asks you to work out the COUNT of users who activated in 2019 (note: a customer is considered activated when he goes through at least one successful deposit and places a bet). How would you do this with SQL code? Are there any other constraints you might consider adding to give a more accurate estimate?

SELECT COUNT(DISTINCT u.playerid)

FROM USER u

INNER JOIN ACTIVITY a

ON u.playerid = a.playerid

WHERE YEAR(u.activation\_date) = 2019

AND YEAR(a.betplaceddate) = 2019;

or with a VIEW you could (Mainly for later)

CREATE or REPLACE VIEW market AS

SELECT \*

FROM USER u

INNER JOIN ACTIVITY a

ON u.playerid = a.playerid

WHERE YEAR(u.activation\_date) = 2019

AND YEAR(a.betplaceddate) = 2019;

SELECT COUNT(DISTINCT playerid)

FROM market;

1. The CRM Manager is happy with the base user size so it’s time to generate the list of users to be sent the email. With the same constraints as above you need to generate a list with the following fields, so that the CRM Manager can start to build his custom segments:
   1. playerid
   2. alias
   3. email
   4. total stake generated in 2019
   5. margin in 2019 (i.e. gross revenue/stake)
   6. last bet placed date
   7. total amount deposited on Sportsbook (approved deposit only)

How would you create this list in SQL?

Using the view I created about:

SELECT m.playerid, m.alias, m.email, SUM(m.stake) as total\_stake, (m.ggr/ m.stake) as ggrm, max(d.betplaceddate), SUM(d.payment\_amount)

FROM market as m

INNER JOIN DEPOSIT AS d

ON m.playerid = d.playerid

WHERE d.payment\_stats =’a’

GROUP BY m.playerid, m.alias, m.email

1. The Marketing VP is concerned that there has been a decrease in Parlay bets (i.e. bets with more than 1 leg) as a % of total bet COUNT. Specifically he wants to know in each month from January 2019 to April 2019 what percentage of Parlay bets we had. Can you provide the SQL code required to answer his questions?

CREATE OR REPLACE VIEW bet AS

SELECT sp.months, SUM(sp.straight)AS straight\_total, SUM(sp.parlay)as parlay\_total

FROM(SELECT DISTINCT wagerid,

SUM(case when betttype = 'straight' then 1 else 0 end)/COUNT(\*) as straight,

SUM(case when betttype = 'parlay' then 1 else 0 end)/COUNT(\*) as parlay,

EXTRACT(month from placed\_date) as months

FROM ACTIVITY GROUP BY placed\_date, wagerid

HAVING placed\_date BETWEEN '2019-01-01' AND '2019-04-30') as sp

GROUP BY months;

SELECT round(straight\_total/(straight\_total+parlay\_total)\*100, 2) as straight\_percent,

round(parlay\_total/(straight\_total+parlay\_total)\*100, 2) as parlay\_percent, months

FROM bet

GROUP by months, straight\_total, parlay\_total

1. The Acquisition Manager is concerned that there has been a decrease in Registration to Activation rate (i.e. many customers registered but they never made a deposit). Specifically he wants to know what Reg-to-Act ratio we have had in each month from January 2019 to April 2019. Can you provide the SQL code required to answer his questions?

CREATE OR REPLACE VIEW activation as

SELECT na.months, SUM(na.straight)AS nonactive\_total, SUM(na.total)as total

FROM(SELECT DISTINCT playerid,

SUM(CASE WHEN activation\_date IS NULL' THEN 1 ELSE 0 end)/COUNT(\*) as straight,

COUNT(bet\_type) as total,

EXTRACT(month from placed\_date) as months

FROM bets GROUP BY placed\_date, playerid

HAVING placed\_date BETWEEN '2019-01-01' AND '2019-04-30') as na

GROUP BY months;

SELECT months,

CONCAT(CAST(round(nonactive\_total/(nonactive\_total+total)\*100)AS text),'/100' ) AS Non\_Active\_Ratio

FROM activation

GROUP by months, nonactive\_total, total