Data Analyst - Case

Please create a document/presentation and be ready to discuss the following case. Note: there is no "right" answer, we're more interested in your thought process and communication skills rather than the answers themselves.

The following case uses a hypothetical scenario combined with fictitious data to evaluate your practical and analytical acumen. Any real or perceived similarities to real world scenarios are purely coincidental.

Question 1

We maintain a table named "spending" in our analytics database. The table contains a record for every purchase that has taken place in our game. It has the following columns:
- player id (the unique identifier for the player who made the purchase)
- spend (the amount of money spent in USD)
- timestamp (the timestamp at which the purchase took place)

Below is an example of what the table contains:

player_id	spend	time_stamp
337079	31.32	10/30/2014 2:38
590056	22.04	10/30/2014 18:43
775561	33.35	10/30/2014 21:07
1010417	3.8	10/31/2014 4:33
547471	9.96	10/31/2014 11:02
366382	31.32	10/31/2014 12:28
440351	5.6	11/1/2014 6:00
276079	38.28	11/1/2014 14:24
322389	24.36	11/4/2014 8:38
191993	9.6	11/4/2014 21:36
355505	15.12	11/5/2014 15:36
478672	27.26	11/6/2014 3:36
318815	16.92	11/6/2014 22:33
.	.	.
.	.	.
.	.	.

Additionally, we classify spenders according to how much money they have spent in the game.

- "minnows" have spent less than $50.00
- "trout" have spent between $50.00 and $99.99
- "salmon" have spent between $100.00 and $499.99
- "tuna" have spent between $500.00 and $999.99
- "whales" have spent $1,000.00 or more

Please write a query using PostgreSQL syntax that will produce the total spend each day by each type of spender. Please provide your query in a text file.

An example of the result is below.

date	minnows	trout	salmon	tuna	whales
10/30/2014	117.27	38.52	697.6	181.17	909.81
10/31/2014	99.21	121.04	638.22	233.64	1000.89
11/1/2014	59.79	30.24	442.3	371.25	1079.1
11/2/2014	111.13	60.48	577.63	0	1606.77
11/3/2014	130.9	104.65	448.63	144.54	1515.69
11/4/2014	101.65	67.56	373.12	0	1309.77
11/5/2014	106.41	82.44	352.35	211.86	696.96
11/6/2014	100.72	47.64	611.32	184.14	894.96
11/7/2014	109.64	110.29	630.17	180.18	1420.65
11/8/2014	124.96	61.2	445.31	233.64	1375.11
11/9/2014	94.99	157.86	438.77	370.26	1497.87
11/10/2014	121.35	122.88	511.56	394.02	1378.08
11/11/2014	118.84	69.58	738.05	64.35	1665.18
.	.	.	.	.	.
.	.	.	.	.	.
.	.	.	.	.	.

Question 2

We track a wide variety of metrics that we use to analyze player behavior inside our games.

Examples of these include average session lengths per day per player, average runs per player at each game level, and average revenue per daily active player. These three metrics are just the tip of iceberg. We collect a tremendous amount of data and use it extensively to evaluate our games.

Please list 10 metrics that you would include in a report to assess player behavior in Subway Surfers.

Question 3

One great way to optimize a game is to conduct experiments. For example, we can change how fast Jake runs or increase the number of coins players receive. We can then use data to assess whether the game changes provided a better or worse player experience.

- What experiments would you want to run in Subway Surfers?
- How would you expect player behavior to change?
- What data would you collect to evaluate the results of the experiment?

Question 4

SYBO's new mobile game, which is a game studio tycoon called *BOSY BOYS*, is planned to be released in six months. The core game, as well as the theme and art design, is already determined but for the meta-game, there are two competing designs that we could not make our mind about. However we need to pick one of them for the final game.

You can imagine these different meta-game designs as two different menu, event and shop designs that engage people in different ways. Therefore these are not core differences that

would change the game dramatically but they would definitely change user behavior in terms of engagement and spending.

During soft launch* we release the game with meta-game A, which we call Bosy Boys A (BBA). After a long enough time we change the meta-game to the other meta-game version which we call Bosy Boys B (BBB).

In the attached .xlsx workbook you can find three worksheets which contain Retention, Revenue and Daily Active Users (DAU) data for the last 28 days of the soft launch. During the first 14 days (Days -14 to -1) BBA is live. On the Day 0 SYBO updates the game to BBB, which is the live meta-game during the last 14 days (Days 0 to 14). Assume that all players are obligatory to update their game in the beginning of Day 0.

Also assume that since the meta-game does not change the game radically, it attracts more or less the same audience despite the change.

Using this data:

1. Draw
    a. DAU by date graph
    b. Revenue by date graph
    c. Retention** by day graph ***
    d. Any other graph that might be required or helpful for your argument in 2)
2. Which meta-game is a better choice for BOSY BOYS? BBA or BBB? Please substantiate your answer by making a case about your decision. We are more interested in your approach and reasoning than the actual answer.

* Mobile game companies usually make soft launches in a limited number of countries to test user response.

** Day N Retention is defined as the percentage of users who return to the game on Day N after installation. E.g., 50% Day 1 retention means 50% of the users return to the game one day after they install the game.

*** Note that in this context, "day" is the number of days after a player installed a game. The retention data includes values out of the soft launch window (e.g., Day 90). This is so since this data is an extrapolation of the true retention data based on a predictive model.