I. Spam Friend Request

Frequency: 4
Mocked? ✔️

SQL Questions:

Q1: For each month in 2020, what is the friend request acceptance rate?

- request_id, accepted, group by month
- countif(accepted), count all

Q2: What percent of users in the United States accepted a friend request last week?

- user_id, accepted_one_request, filter by US
- ratio: user accepted at least one / all user in the us

```
Table:
   users
Columns:
   user_id            INT      The unique identifier of the Facebook user
   country            STRING   Country of the Facebook user
   facebook_age       INT      Days since the user signed up for Facebook

Sample Rows:
user_id  | country | facebook_age
-------------------------------------
  1001    |  'US'   |    120
  1002    |  'CA'   |    400
  1003    |  'MX'   |    6


Table:
   friend_requests
Columns:
   date               STRING   The date a friend request was sent
   sender_id          INT      The user that sent a friend request
   receiver_id        INT      The user that received a friend request
   result             INT      0 if no response, 1 if accepted, 2 if
rejected

Sample Rows:
date        | sender_id |  receiver_id  | result
-------------------------------------------------
```

```
'2020-01-01' |   1001    |     9991      | 0
'2020-01-01' |   1001    |     1032      | 1
'2020-01-01' |   1002    |     5555      | 2
'2020-01-01' |   1002    |     1234      | 1
```

Applied Data

1. How to detect spam requests?
    a. Always start with the goal: decrease the bad experience created by spam friend request so that users could have a consistent and healthy experience
    b. Ground truth: define what is a spam?
        i. User feedback: Deleted the requests
        ii. Identity, connections, interactions,
    c. Labeled data, how to design the algorithm?
2. What would be the features to include in your algorithm?
    a. Elaborated by above
3. How do you apply this algorithm?
    a. Block completely
    b. Down-rank the requests
    c. Add a flag to the spam friend requests
4. How would you measure the launch of the algorithm?
    a. Primary metric: acceptance rate increases
    b. Secondary metric: number of connections per user,
    c. Counter metric: product DAU because of the increase in the connections

Product Interpretation:

5. # of friend requests dropped by 10%
    a. Low-hanging fruit: verify this change is unexpected
        i. Competitor,
        ii. seasonality: YoY,
        iii. platform-wise: other metrics?
        iv. technical issues
    b. Less exposure to send friend requests?
        i. Impressions of friend request button / user
    c. The recommendation makes no sense
        i. CTR rate of the friend requests
    d. Users are genuinely less interested in sending friend requests
        i. # of searches for people / user
6. Product manager comes to you and suggests that *People you may know* algorithm is the reason why the number dropped, how do you validate that?
    a. DiD of the CTR
    b. A/B test again

7. New algorithms to detect spam friend requests, how would you go about measuring if the algorithm works better?
    a. Primary metric: Average CTR of the top k results
    b. Secondary metric: Acceptance rate, Initiated first conversation
    c. Counter metric: product DAU vs messenger DAU

III. Dwell Time

Frequency: 4 times
Mocked?

Product Interpretation:

⚠ Q1: When a user" opens the Facebook page" and "closes the Facebook page", we say it is a "session" for an active user. For example, an active user can log in and log out for 10 times a day. Then, we say this use has 10 sessions on that day. How do you measure which sessions are valuable for user?
Valuable: performed meaningful actions to the users
Users:
content creator: high throughput, low quality, high quality, low throughput

- # of sessions with post > 10 / user
- # of sessions with posts reactions / comments / > 100 / user

consumer - deeply engage vs surfer

- Passive: # sessions with impressions > x
- Active: # sessions with engagement > x
- With specific intention: # active search / check birthday / words with friends

Q2: There're two major types of users, one who loggin and just browse, one who deeply engage within each session.

Technical Analysis: SQL Question (onsite)

```
User_id | event_time | event_name | session id
                     open menu
                     close menu
Table1: user-sessions
date | user_id | start_time | end_time | end_reason
-----------------------------------------------------------------
'2020-09-01' | 1 | 1964783746 | 1964783924 | 'close-app'
'2020-09-01' | 1 | 1964783528 | 1964783809 | 'crash'
'2020-09-02' | 2 | 1964783123 | 1964783345 | 'close-app'
'2020-09-02' | 3 | 1964783252 | 1964783658 | 'crash'

Table2: dim_all_users
user_id | country | is-active
---------------------------------
1 | 'US' | TRUE
2 | 'US' | FALSE
```

```
3| 'CANADA' | TRUE
```

1. What is the average time (in seconds) each user stays on Facebook on specific date?
    a. user_id, avg_dwell
2. What is the percentage of active users who crashed when using Facebook from the US area on a specific date?
    a. user_id, crashed_at_least_once, country
    b. count(crashed) / count(active users)
3. Calculate the average dwell time in seconds across all sessions (i.e. return one number)? Dwell time is the length of time between opening and closing the menu.
    a. avg(end_time - start_time) using lag
4. Get the percentage of all sessions that have both nav_menu_open and nav_menu_close?
    a. session_id, open, close, by self_join
    b. countif(open is not null and close is not null)/ count(session_id)
5. Lets say we want to account for missing events by setting the dwell time to 60 seconds whenever a nav_menu_close event is missing. Can you write a query to re-calculate the new average dwell time when we default to 60 seconds of dwell time whenever nav_menu_close is missing?
    a. session_id, open, close by self_join
    b. avg(COALESCE(close-open, 60))
6. Calculate the average gap between each session.
7. Calculate the app bounce rate, from app_1 -> app_2 -> app_1

V. Instagram

Frequency: 1
Mocked? No

Technical Analysis: SQL Question

Product Interpretation: Product Sense

1. Instagram's Android users engagement is really low, how would you go about analyzing it?
2. If you were the DS for stories, PM comes to you and mentions that the number of reads for Instagram Stories is lower than Facebook Stories.
3. Given the hypothesis you wanted to analyze is true, how would you go about changing it?
4. How would you run an A/B test around it?

Facebook stories vs Instagram Stories

1. Facebook stories' number of views < Instagram's, why?
    a. Product level
        i. The number of DAU facebook < instagram
    b. Feature level
        i. Content creator
            1. Fewer content creators
            2. Less motivation to share on facebook: a real-life social network, or less followers than instagram
                a. Stories being viewed by friends vs public ratio
                b. Follower friends / public ratio
            3. Tools are not conducive compared to instagram
        ii. Consumption
            1. Less exposure
            2. Less relevant, exciting stories
            3. Genuinely less interested

Applied Data: Modeling

Quantitative Analysis: Statistics and Probability

VII. Messenger, Payment

Technical Analysis: SQL Question

Product Interpretation: Product Sense

1. What is Messenger's value to Facebook?
    a. Product level:
        i. User engagement, longer term strategy: Private-message: time spent
        ii. Revenue: payment, auto-reply robots for business, ads
    b. Ecosystem:
        i. More dependencies
        ii. More time spent in the system
        iii. More data collected about the user
2. How to define DAU of Messenger?
    a. 30-day average of daily login users who performed one of the meaningful actions
        i. Engaged in a conversation
        ii. Sent payment
        iii. Clicked into an ads
3. If DAU of Messenger increase, but % using decreased, why?
    a. Number of new users joining, fewer contacts,
    b. % = users with engagement / DAU
        i. Less exposure to engagement
        ii. Less relevant notifications
        iii. Users are genuinely less interested in chatting in messenger (longer-term issue) check with competitors

Applied Data: Modeling

Quantitative Analysis: Statistics and Probability

distribution of number of transactions, mean, median,
distinguish small business and suspicous account

IX. Spam Account / Posts / News

Frequency:
Mocked?

Technical Analysis: SQL Question

Product Interpretation: Product Sense

1.
2.

Applied Data: Modeling

3. Facebook designed a new spam filter to downrank spam posts, how do you evaluate the success?

Post 1, ⚠

4. What's the impact of fake news on Facebook, you'll have 8 hours to analyze this, what would you look at and what would you do? (No historical records, no user report, no NLP scoring, no A/B testing)
5. How would you define the impact?
6. If you were asked to develop a model around this, what variables would you use?

The following features

```
1. Account registration time
2. Amount of purchase
3. Transaction Log - history of ad purchases
4. IP address log
5. Ad image / URL for the current purchase
6. Change of address / email / phone
7. # of friends
8. Account access
9. Call account holder
```

Quantitative Analysis: Statistics and Probability

⚠ Spam account is 1% of total user, the spam detection model accuracy is 95%

1. What is the probability that the user is a spam if the model predicts the account is a spam?

2. Out of 100k users, how many of the users are spam accounts?
3. What is the variance | standard deviation of the prediction out of 100k users?

XI. Ads

Frequency:
Mocked?

Technical Analysis: SQL Question

Product Interpretation: Product Sense

1. The Facebook Integrity team wants to test the precision of new enforcement that is meant to target spammers on Facebook. This enforcement blocks all messages that happen between users who don't share any connections (mutual friends). The team wants you to design the experiment in a way that will provide conclusive results.
2. How would you design and set up this experiment?
3. What information, if any, would you need to properly determine the sample size and length of the experiment?
4. Which metrics would you plan to track in order to define the performance of the enforcement? Why did you choose these metrics?
5. How would you identify Fake Accounts on Facebook/Instagram?

Branded Ads and Direct Response Ads
Q1. What is the difference between branded ads and direct response? (Context: advertisers have separate budgets)
Q1.1 How do you define branded vs direct response ads
Q2. How do you evaluate the hypothesis: Facebook is not good on brand advertising. Advertisers are spending less budget on it.
Q3. What if branded ads revenue decreased?

Video Length
Q1. Shall we launch longer video if the longer the video, the higher the chances of users clicking into the video. (overall engagement of videos vs immediate ads revenue; longer-term metrics like engagement, DAU)
Q2. Facebook launched video ads products a few years ago. How do you know if it's helping advertisers?

Not relevant:

有些users会用偷来的信用卡在FB上买广告。以下有9个feature，如果你只能选择3个feature来predict 一个user有没有用偷来的信用卡，你会选哪三个。

1. Account registration time
2. Amount of purchase
3. Transaction Log - history of ad purchases. check 1point3acres for more.

4. IP address log
5. Ad image / URL for the current purchase
6. Change of address / email / phone
7. # of friends
8. Account access-baidu 1point3acres
9. Call account holder

这题没有正确答案。因为不同的fraud type会用到不同的feature。最后我们把每个feature都讨论了一边。这一轮基本只考business intuition；面试官会给很多hints，因为很多面试者没有biz context。

Applied Data: Modeling

Quantitative Analysis: Statistics and Probability

XIII. News Feed

Frequency:
Mocked?

Technical Analysis: SQL Question

```
# Table 1: recorded all the comments posted by users
user_id | ds | comments
ds --# date
comments -- # of comments posted by the user
# Table 2: recorded all the user information and country
user_id | country -- country where the user register at
```

Q1. return the total number of posts per user in May 2020
Q2. return the percentage of users that posted comments in May 2020

Product Interpretation: Product Sense

Q1: What if the product team wants to change the layout of news feed, such that more feeds could be seen, i.e. the size of a single post shrinks - you'll need to click 'See more' to see the full post. Do you think it's a good idea?
Q2: How to measure if this is a good product change?
Q3: If two countries have varying responses to the product change, why do you think it's the case and how do you measure it?

Recommend other groups
Q1: How to decide on the best ratio of 'Recommend other groups' post vs other feeds (especially ads) in the news feed? (Find the right balance, trade-off, baseline -> A/B test!)
Q2: Brainstorm ideas to increase the number of comments for the Facebook group posts

Public posts vs private posts
Q1. How to decide on the best ratio of public posts vs private post?

Applied Data: Modeling

Potential answer
⚠ Q1. How do you measure the negative impact of reshare on facebook? (potentially bad news reshare, also popular posts being prioritized more like top news, not)

Quantitative Analysis: Statistics and Probability

XV. Best Friend on Facebook

Frequency:
Mocked?

Technical Analysis: SQL Question

Product Interpretation: Product Sense

Applied Data: Modeling

Q1: How do you determine a user's best friend on facebook?
Q2: What's the most important variable in determining the relationship?

Quantitative Analysis: Statistics and Probability

XVII. Notification feature

Frequency:
Mocked?

Technical Analysis: SQL Question

Q1. Can you write a query to compute the median number of notifications users see each day?
Q2. using the same table given in the last question, can you write a query to compute the daily dismissal rate for each notification type?
Q3. Another table is given with columns: date, time, user_id, event_name. Can you write a query to fetch the percentage of users who saw a 'suggested_friend' notification on each date who also added a friend on the same day they saw the notification?

Product Interpretation: Product Sense

Q1. How would you design the notification feature?
Q2. What is your hypothesis? How do you verify the hypothesis?

Q1. Say you are on the team, What kinds of metrics should we be measuring to understand the health of the notifications ecosystem? How should we define these metrics?
Q2. Looking at our data, we know that there are instances where users ignore notifications. What hypothesis do you have to explain why this occurs?
Q3. What are the pros and cons of each of the three spamminess metrics we discussed (notifications/user, notification dismissal rate, notification ignore rate)? How would you use these metrics, and what other metrics would you look at?
Q4. How would you interpret the metric you computed in the previous question and how is it different from a 'conversion rate' metric?

Applied Data: Modeling

Quantitative Analysis: Statistics and Probability

XVIII. Brand Marketing Measurement

Frequency:
Mocked?

Technical Analysis: SQL Question

Product Interpretation: Product Sense

What are the metrics that advertisers care the most? Of course, how do you make them think that FB advertising makes sense?

Applied Data: Modeling

Quantitative Analysis: Statistics and Probability

XX. Parents on Facebook

Frequency:
Mocked?

Technical Analysis: SQL Question

Product Interpretation: Product Sense

Very good answer
Q1: What is the impact of parent's presence on Facebook?
Q2: How to measure the engagement of the kids/teenagers?
Q3: When designing the experiment, how would you randomize to avoid biases, e.g. factor other than parents joining.
Q4:

Applied Data: Modeling

Q1: How do you measure the impact of parents' presence on Facebook?
Q2: How do you improve, knowing that there's an impact on Facebook?
Q3:

Quantitative Analysis: Statistics and Probability

XXI. Portal

Frequency:
Mocked?

Technical Analysis: SQL Question

```
# Table 1 video_calls:
caller| recpient| ds| call_id| duration
# Table 2 fb_dau:
user_id| DAU_flag| ds| country
```

Q1: On 2020-01-01 how many people initiated multiple calls?
Q2: % of DAU used the video calls function by each country on 2020-01-01?

Product Interpretation: Product Sense

A sample answer
Q1. Facebook messenger has video calling features, they have a product called Portal, currently selling at around $200.
Q1.1 How do you determine the number of target buyers; who would buy?
Q1.2 How do you find out who has the disposable income to purchase Portal?
Q1.3 How do you find out if Portal has met the needs of purchasers?
Q2. Facebook plans to sell in Europe but to get enough market share, may need to price at $20 below cost, is it worth it? Just focus on the video call feature.
Q2.1 What's the potential impact of Portal to Facebook overall? (ecosystem dependency)
Q2.2 How do we decide if we want to launch a 1 for 1 promotion?

Follow up questions:
how to define a potential buyer? (distribution of number of calls per day/ per week at a fixed location)
How to measure the increase of DAU is due to the launch of the product?

Applied Data: Modeling

Quantitative Analysis: Statistics and Probability

XXII. Instagram switch account

Frequency:
Mocked?

Technical Analysis: SQL Question

Product Interpretation: Product Sense

Applied Data: Modeling

Quantitative Analysis: Statistics and Probability

XXIV. Facebook Memory Page

Frequency:
Mocked?

Technical Analysis: SQL Question

Product Interpretation: Product Sense

Applied Data: Modeling

Quantitative Analysis: Statistics and Probability

XXV. Health of Facebook Group

Frequency:
Mocked?

Technical Analysis: SQL Question

Product Interpretation: Product Sense

Applied Data: Modeling

Quantitative Analysis: Statistics and Probability

XXVI. Instagram switch account

Frequency:
Mocked?

Technical Analysis: SQL Question

Product Interpretation: Product Sense

Applied Data: Modeling

Quantitative Analysis: Statistics and Probability

XXVII. Facebook SMS verification

Frequency:
Mocked?

Technical Analysis: SQL Question

FB -(send verification code)-> mobile carriers -(send as sms)-> user -(confirm)-> FB。 FB有发给carrier的log和user的confirm的记录（见下），但从carrier到user是黑盒。FB一天可以向carrier发送多次请求，但用户一天之用确认一次。

Q1. How many requests have we sent to the carriers?

```
sms_send
| ds (date) | carrier | country | phone num | event_type |
event_type：发送信息的类型，包括"confirm", "recovery", etc

confirm
| ds | phone |
```

Product Interpretation: Product Sense

Q1. If the number of confirmation dropped by x%, what could we do about it?
Q2. If the number of sent requests stands constant, the no. of confirmation dropped by x%, what could've gone wrong?
Q3. If we could confirm that the drop was due to the carrier, how do we find out which carrier it was? (confirmation rate WoW by carrier)
Q4. If all the carriers had downtime, FB needs to contact all 5 carrier, how should we choose?

Applied Data: Modeling

Quantitative Analysis: Statistics and Probability

XXVIII. Workspace

Frequency:
Mocked?

Technical Analysis: SQL Question

```
Region | country | company_id | no_of_onboarded_employee
```

Q1. Calculate the average number of onboarded employees per country per region

Product Interpretation: Product Sense

Q1. We recently identify that the average number of onboarded are decreasing recently, what are the reasons you can think of?
Q2. We are recently thinking about publishing a DO NOT DISTURB function, what are you going to do to determine if you are going to publish this feature? (linked to Q1, users are less likely to be onboarded if disturbed constantly)

Applied Data: Modeling

Quantitative Analysis: Statistics and Probability

XXIX. Restaurant

Frequency:
Mocked?

Technical Analysis: SQL Question

Product Interpretation: Product Sense

Applied Data: Modeling

Group restaurant recommendation, build a new feature

Quantitative Analysis: Statistics and Probability

XXX. Facebook Shop

Frequency: 2
Mocked?

Technical Analysis: SQL Question

```
date | order_id | sender | timestamp, sender = 1: seller, sender=0: buyer
```

Q1. how many orders have message from buyer on date xxxx.
Q2. how many orders have messages both from buyer and seller.
Q3. how many orders with last message sent from seller.

Product Interpretation: Product Sense

Q1. How to determine if a seller is bad or good?
Q2. Follow-up: how to determine if an auto-reply is good or bad, how to quantify that?

Applied Data: Modeling

Quantitative Analysis: Statistics and Probability

XXXI. Question Template

Frequency:
Mocked?

Technical Analysis: SQL Question

Product Interpretation: Product Sense

Applied Data: Modeling

Quantitative Analysis: Statistics and Probability

XXXII. Question Template

Frequency:
Mocked?

Technical Analysis: SQL Question

Product Interpretation: Product Sense

Applied Data: Modeling

Quantitative Analysis: Statistics and Probability

DAP: Number of daily active people, measured every day.
Likes: Total number of daily likes given in Facebook, measured every day.
Likes/DAP (likes per DAP): Measure of engagement.

Q1: Likes per DAP is 10. What do you expect the median likes per user to be? > 10, < 10, ~10? Why?
How does the distribution of likes look like?

Q2: This week, likes per DAP is up 10% y/y. Why do you think that is? */

Facebook (including mobile app and in-app browser)
Messenger
Instagram
Portal-branded devices
Oculus Products
Facebook Shops
Facebook Business Tools

Audience Network
Spark AR Studio

AB Test Questions

1. AA test vs AB test?