Note: 请大家不要share，违者必究

Week 1:

W: SQL - questions using “joins”
F: Common probability questions at interviews (playing cards, dice…)
Su: SQL - questions using “joins” and others part II

预习资料：

Read all tutorials below
https://www.w3schools.com/sql/
https://stackoverflow.com/questions/5706437/whats-the-difference-between-inner-join-left-join-right-join-and-full-join
http://www.math-only-math.com/playing-cards-probability.html
http://www.math-shortcut-tricks.com/probability-problem-on-dice/
https://databricks.com/blog/2015/07/15/introducing-window-functions-in-spark-sql.html

Questions:

(W)

1. We are studying ecommerce advertisers on Wechat over a certain time period (say a week). The time period does not matter for this problem. You are given 2 Tables:
adv_info: advertiser_id, ad_id, spend (primary key: ad_id)
ad_info: ad_id, user_id, price (primary key: ad_id, user_id)

Adv_info table contains information on advertisers. Advertiser_id is id of advertiser
; ad_id is id of an ad being run by advertiser; spend is amount of money in $ that advertiser pays wechat for ad-id to show it to Wechat users; price is the revenue for a specific ad.

Q1: What would be the average advertiser spend on wechat? Your query should return a single number.
Q2a: find advertisers with at least one conversion
Q2b: what % of advertisers have at least one converted user.
Q3: We want to come up with an advertiser level metric that quantifies how well wechat advertising is working for advertisers. This should be based on the above 2 tables.
After coming up with this metric, We want to compute it for each advertiser.

(F)

1. Given a deck of 52 cards, what's the probability of choosing two cards that are not in the same suit(花色）and not a pair（对子)?
2. 一个人去赌场，花5刀玩一个游戏，扔两次股子，如果和为6，他赢21刀，其他就什么也没有问，这个游戏是偏向casino还是player的? Follow up: 如果现在有一个策略，这个人一直玩，直到第一次赢，然后走。问第一次赢需要“平均”玩几次？

3. 在一个班级里100人一起扔每个人手里的一个硬币，问一起扔后得到50个head50个tail的概率是多少？
4. 已知一对夫妇有两个孩子，其中一个是男孩，请问另一个是男孩的概率？

(Su)
给一张表描述Line每天聊天数量的记录(n_msg = # of messages)
date | user1 | user2 | n_msg
（1) 从这个表我们可以知道些什么信息?
（2）写个query得到某一天用户发消息朋友数量的distribution，就是output出两列，X: number of unique contacts for each user; Y: number of user with this many contacts。你觉得这个distribution会长什么样子，为什么？
（3）写个query找到每个user发信息最多(所有消息总和数）的top partner