Assignment 3 Complex queries

Rules:

All the questions and the sample result sets in this assignment are using postgres database, and you can import data from **shenzhen_metro.sql**

- Your result set, especially the data type and the order of each column, must strictly follow the description and the sample result set in each question.
- You need submit ".sql" files for these five questions.
- The name of each ".sql" file should be q1, q2, q3, q4, q5 respectively to represent these five questions.
- Do not forget to add ';' in the end of each query.
- Do not compress them into a folder, please submit them directly.
- Please submit those queries into **sakai website** as soon as possible, so that you can get chance to receive feedback before deadline. After the deadline, we will check the assignment automatically by a script and then given your grade, at that time, any argument about your grade of this assignment will not be accepted.

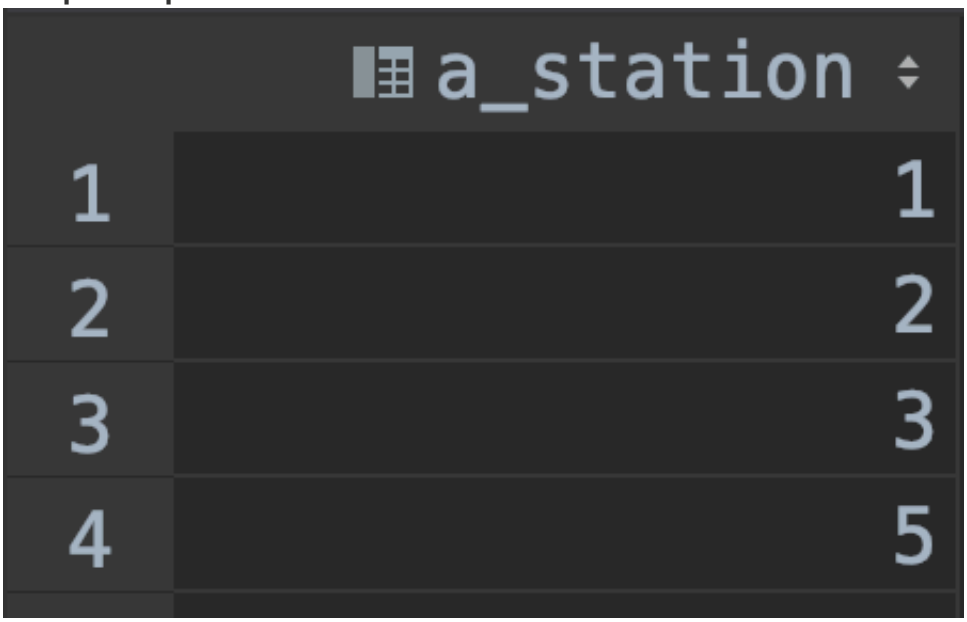
Description:

As a student at the SUSTech, we often use the Shenzhen Metro and buses. In this assignment, we use the data of Shenzhen subway and bus for exercise.

Problem 1

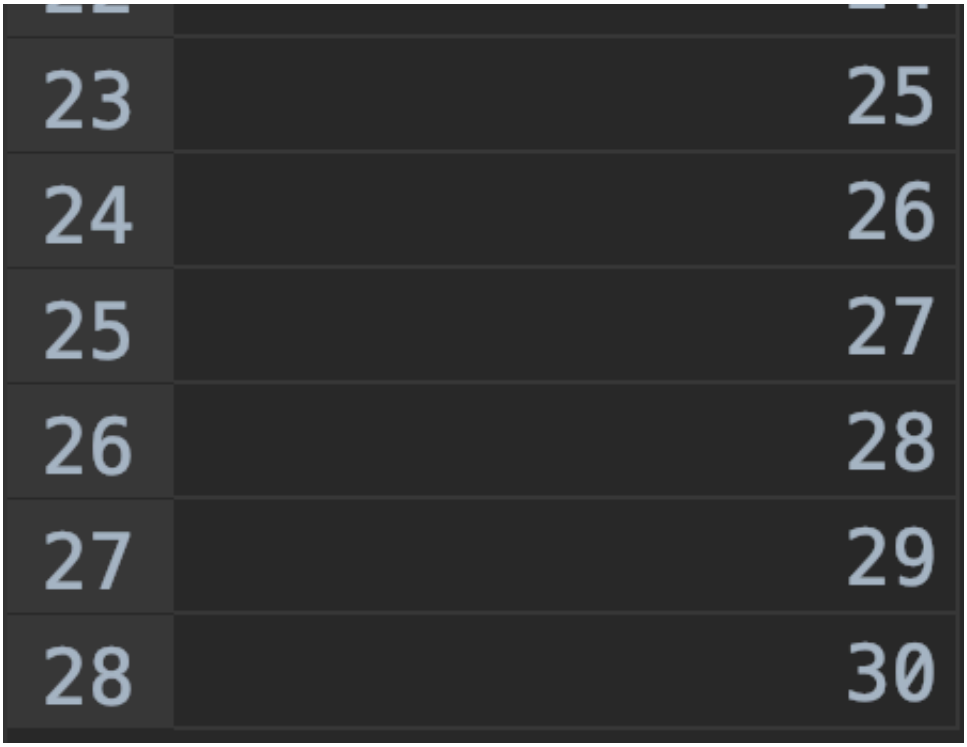
Which stations in Line 1 are not in Line 2, please output the id of those stations in ascending order
请输出一号线上的哪些站不在二号线上,请升序输出车站id?

Sample Output:



	a_station
1	1
2	2
3	3
4	5

5	6
6	7
7	8
8	9
9	10
10	11
11	12
12	13
13	14
14	16
15	17
16	18
17	19
18	20
19	21
20	22
21	23
22	24



23	25
24	26
25	27
26	28
27	29
28	30

Problem 2

**The stations on Line 1 distributed in which district? How many stations are there in each district? Please output the district names, the number of stations, and the ranking.

Order of result set is ignored by Testing script

**一号线上的车站都分布在哪些区？这些区每个有多少个属于1号线的车站？请输出这些区名，车站数量，以及排名。

测试时忽略排序问题

Sample Output:

	district	number	rank
1	Luohu	4	1
2	Bao'an	8	2
3	Futian	9	3
4	Nanshan	9	3

Problem 3

Please output how many subway lines pass through each district and the ranking of the number of subway lines in each district.

Order of result set is ignored by Testing script. and the null district should not take into consideration.

请输出每个区有多少地铁线路经过，以及每个区地铁线路数量的排名。。

测试时忽略排序问题, 数据库中字段区（district）为null的站点不纳入考虑

Sample Output:

	district	number	rank
1	Luohu	7	1
2	Futian	7	1
3	Nanshan	6	3
4	Bao'an	3	4
5	Longhua	2	5
6	Longgang	2	5

Problem 4

Please output the subway stations with more than or equal to 10 bus stops around the subway station in each line, sorted by 1. ascending order of line id, 2. descending order according to the count of bus stops, 3. descending order of station id. Your result set only return 10 rows from the 16th row (You can use limit 10 offset 15)
请输出每条线路中地铁站周边公交站多于或等于10的地铁车站，排序依据 1. line id 的升序，2. 公交车站数的降序， 3. station id 的降序。返回结果集从第16行开始保留10行

Sample Output:

10 rows

	line_id	station_id	cnt	rank
1	1	1	44	7
2	1	16	45	6
3	1	2	48	5
4	1	18	49	4
5	1	17	52	3
6	1	15	55	2
7	1	12	59	1
8	2	62	10	16
9	2	59	10	16
10	2	63	11	13

Problem 5

Once upon a time, a new intern comes to Shenzhen metro department to design new stations. As a biology under-graduate, the intern is asked to simply give names to stations instead of designing routes. But the master of department is a strange guy who dislikes the station starts with the same character. For example, if some station starts with '深' such like '深云' and another station named '深大', then the frequency of this character is 2 (in fact considering all stations, '深' presents 7 times).

However, human beings' ability has the limits and thus the intern cannot find new names with never-used start character. So the master relaxed the limitation and **if the name doesn't start with the character that shows up the most times, then the name is valid**. However, the limitation asks for **'the most frequency for each district'**. So, although the character '宝' is not valid in Bao'an district, it's valid in Nanshan district.

Even worse, **some district has the same frequency for some characters**. For example, '深', '华', and '香' are all presents 4 time in Futian district, and all of them are not valid.

The intern has this task daily and nightly in his thought. Finally, on a stellar scintillation night, he had a dream about a short paragraph of SQL code that can list districts, all the most frequency words for each district, and their frequencies. On the next day, as his best friend, you are asked to write this code for him.

Task: find the most frequency starting characters of stations for each district. For example, in Nanshan district, you should find all the stations in Nanshan district. And count the first character's present time, and get the highest frequency characters' district, character (chr), and present time (cnt).

Hint 1: characters with the same pinyin but not with the same form doesn't same.

Hint 2: the null district should not take into consideration.

Order of result set is ignored by Testing script

某天，小A在深圳地铁部门实习，负责为新地铁站起名。然而，当前同样名字的地铁站实在太多了，而找到完全没用过的名字又很困难，小A只想避免开头第一个字出现很多次的名字。例如，深云和深大都以“深”开头，于是，“深”不能作为新名字的开头。

但人类的能力是有极限的，于是小A打算只去掉出现最多次的那个开头字，在当前情况下，只有“深”不符合要求。但是，他还打算对每个区找到该区出现最多的开头字，例如宝安区为“宝”，南山区为“深”。要注意，有的区可能包含多个同样次数的字，例如福田区的“深”、“华”、“香”均作为地铁站开头字出现了四次，而这些字都是不合法的。

小A日思夜想终于有天梦里找到了一段SQL语句，来帮他找到每个区内出现最多次的开头字。而你是小A最好的朋友，于是他拜托你写出这段SQL语句。

任务：为每个区找到车站名字中出现次数最多的开头字（可能有多个）。例如，对南山区需要找到所有位于南山区的地铁站，并统计名字中第一个字出现的次数，并将最多次数对应的区（district）、字（chr）、次数（cnt）输出。

提示一：同样拼音但字形不同的字不认为是同样的字。

提示二：数据库中字段区（district）为null的站点不纳入考虑。

测试时忽略排序问题

Sample Output:

district	chr	cnt
Bao'an	宝	4
Futian	福	4
Futian	华	4
Futian	香	4
Longgang	大	2
Longhua	龙	2
Longhua	民	2
Luohu	红	2
Nanshan	深	5