1. (2 points) You are a data analyst in the wealth department of a bank, and your supervisor asks you to analyze why the customer satisfaction of all financial advisors in Canada drops a lot in the last year. Here are 3 questions:
   a. What questions will you ask yourself before collecting data?

1. What are the most prominent financial accounts opened with said bank?
2. Any new rules or regulations applied to accounts?
3. Any accounts with recently raised fees or interest rates (was there a promotional offer that expired)?
4. What are all the bank locations?

   b. What data will you collect to kick off the analysis?

I would review all the customer satisfaction reviews and connect all the bad reviews with the bank location. After collecting the locations of all the banks, I would take the initiative to understand the demographics of the bank's location that had the bad customer satisfaction review. Understand the class of people that are giving bad customer satisfaction reviews by knowing their occupation. Once we know more about the person and their unpleasant experience, the easier it is to accommodate.

   c. What type(s) of data analysis do you need?

The data analysis needed would be the "Diagnostic Analysis" analysis, to help find the root cause of why customer satisfaction has dropped a lot in the last year. It will identify behavior patterns in the company that may need work on to help boost customer satisfaction. Adding "Prescriptive Analysis" may be good as well. If the Diagnostic Analysis concludes that customer satisfaction has been dropping due to the market being bearish, "Predictive Analysis" can conclude that at some time in the future, when the market turns bullish, customer satisfaction will go back up.

2. Which columns will not be useful for predicting whether a passenger will survive?

The PassengerId column will have no effect on predicting survival. The Ticket number as well, but only if the ticket number's value is random and there is no association with the ticket number and cabin number for example. The Name could have links to nationality and technically could have some significance, however since the dataset is small, there is high doubt in it having any significance at all.

```python
import pandas as pd
import os

df = pd.read_csv(os.path.join(os.path.abspath(''),'train.csv'))

pd.set_option("display.min_rows", df.shape[0]+1)
pd.set_option('display.max_rows', df.shape[0]+1)

print(df.head(5))
print(df.describe())
```

[3] ✓ 0.0s

```
   PassengerId  Survived  Pclass  \
0            1         0       3
1            2         1       1
2            3         1       3
3            4         1       1
4            5         0       3

                                                Name     Sex   Age  SibSp  \
0                            Braund, Mr. Owen Harris    male  22.0      1
1  Cumings, Mrs. John Bradley (Florence Briggs Th...  female  38.0      1
2                             Heikkinen, Miss. Laina  female  26.0      0
3       Futrelle, Mrs. Jacques Heath (Lily May Peel)  female  35.0      1
4                           Allen, Mr. William Henry    male  35.0      0

   Parch            Ticket     Fare Cabin Embarked
0      0         A/5 21171   7.2500   NaN        S
1      0          PC 17599  71.2833   C85        C
2      0  STON/O2. 3101282   7.9250   NaN        S
3      0            113803  53.1000  C123        S
4      0            373450   8.0500   NaN        S
       PassengerId    Survived      Pclass         Age       SibSp  \
count   891.000000  891.000000  891.000000  714.000000  891.000000
mean    446.000000    0.383838    2.308642   29.699118    0.523008
std     257.353842    0.486592    0.836071   14.526497    1.102743
min       1.000000    0.000000    1.000000    0.420000    0.000000
25%     223.500000    0.000000    2.000000   20.125000    0.000000
50%     446.000000    0.000000    3.000000   28.000000    0.000000
75%     668.500000    1.000000    3.000000   38.000000    1.000000
max     891.000000    1.000000    3.000000   80.000000    8.000000

            Parch        Fare
count  891.000000  891.000000
mean     0.381594   32.204208
std      0.806057   49.693429
min      0.000000    0.000000
25%      0.000000    7.910400
50%      0.000000   14.454200
75%      0.000000   31.000000
max      6.000000  512.329200
```