1. **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:**

* **What questions will you ask yourself before collecting data?**

There are a few key details that need to be ironed out before beginning to collect the customer satisfaction data. The first being which information would I need to gather from the customers. This is a complicated issue because it would need to cover a wide range of details. For example, would any personal information be needed from these people? If so, which pieces of personal information would be most helpful in this investigation? Contact details, employment details and financial details may be needed to accurately determine the drop in quality of the financial advisors.

Next, which information would be needed that relates to the financial advice the advisors are giving. This information is crucial to determine why customers are dissatisfied and perhaps when it is cross-referenced with their personal information, it could shed some light into what is going on. Speaking to a financial advisor and gathering information from them may be beneficial in this situation as they can show their processes, how they determine their decisions for customers, and most importantly, what information they gather from the customer to aid them in their processes.

In addition to determining which information we would need, concerns around the method of obtaining this data and the details surrounding the storing of this data is important as well to ensure no laws are broken and nobody’s personal information is leaked.

It is also important to note that data cleaning might be needed after the data collection phase and what methods we should apply to have the cleanest data possible.

* **What data will you collect to kick off the analysis?**

A variety of data will be needed to accurately determine the satisfaction drop from the customers. First, I would gather some information from the customers. Their name, age, income, and the reasons they sought out financial advice. I would ask them what problems and issues they had with the service. This last question is important as it would narrow down the problems people are having with the services.

Next, it is important to also gather information from the advisors. Which of the clients from the above data gathering they provided services to, what services they provided to them and how they determined these decisions for their clients.

Once this data is collected, the data may need to be cleaned. It is possible that referencing which financial advisor helped which client might be a problem as the financial advisors might not remember helping a specific client. This is an area of the data that I can see might be a problem collecting but it is probable that there are records of this stored somewhere that may be accessed to gather/clean the overall data.

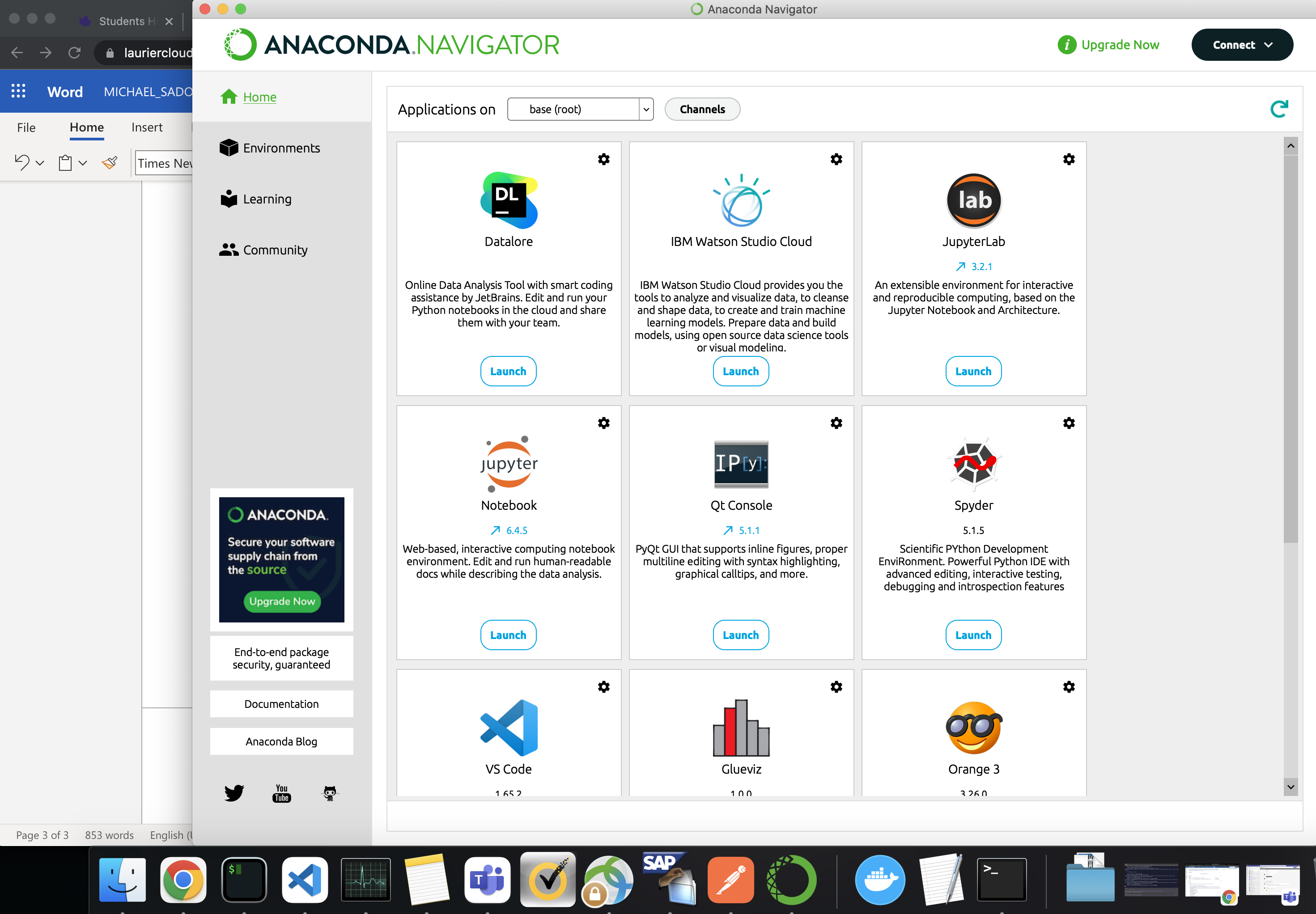
Once this data is gathered and cleaned, further data collection may be needed before a complete analysis is possible and a root cause is determined. This information is a good start and may reveal other problems that would require additional details to be analyzed. For example, if all the advisors are advocating for a specific stock or savings plan, it would be beneficial to find out which of these they are pushing to the clients, collect more data and determine if there is a problem with the specific service.

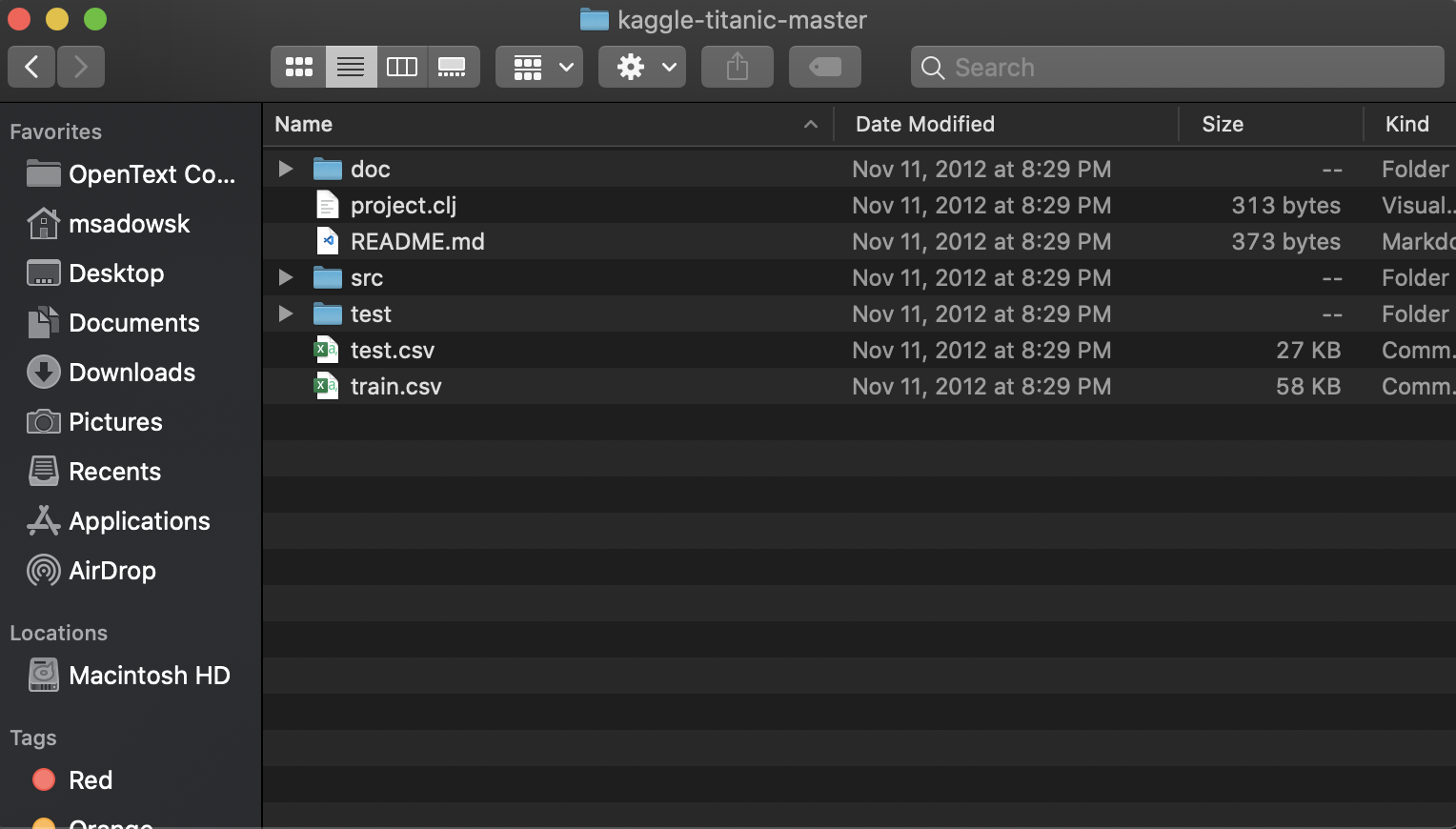
* **What type(s) of data analysis do you need?**

In this scenario, the type of analysis needed would be a combination of diagnostic and prescriptive analysis. The other 2 (descriptive/predictive) do not really apply here.

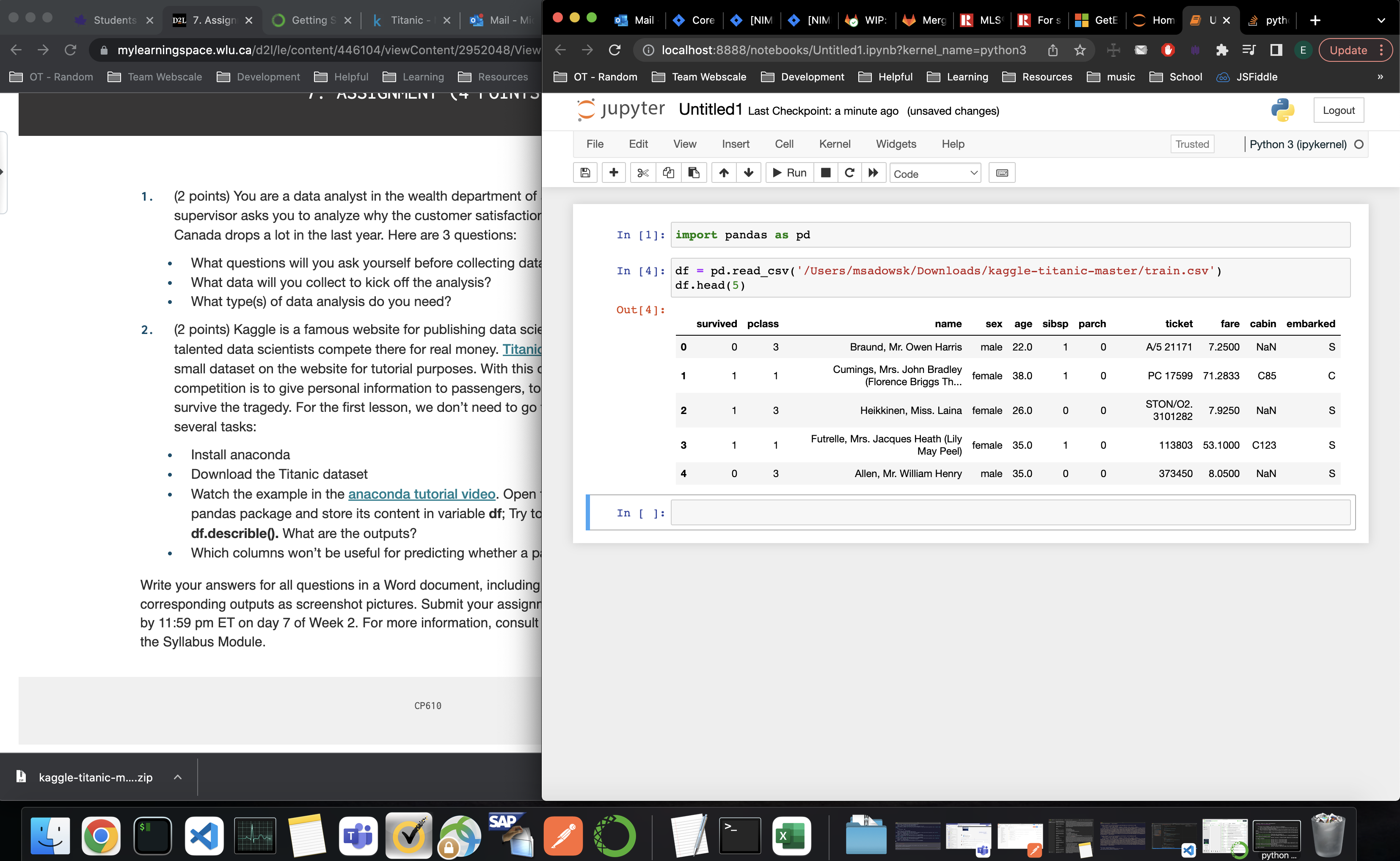
Diagnostic analysis will help determine a root cause for the issues and why customers are providing negative feedback towards the financial advisors. Finding a pattern here is important as the problem describes this is occurring for all their financial advisors, which indicates there is an underlying pattern among the financial advisors.

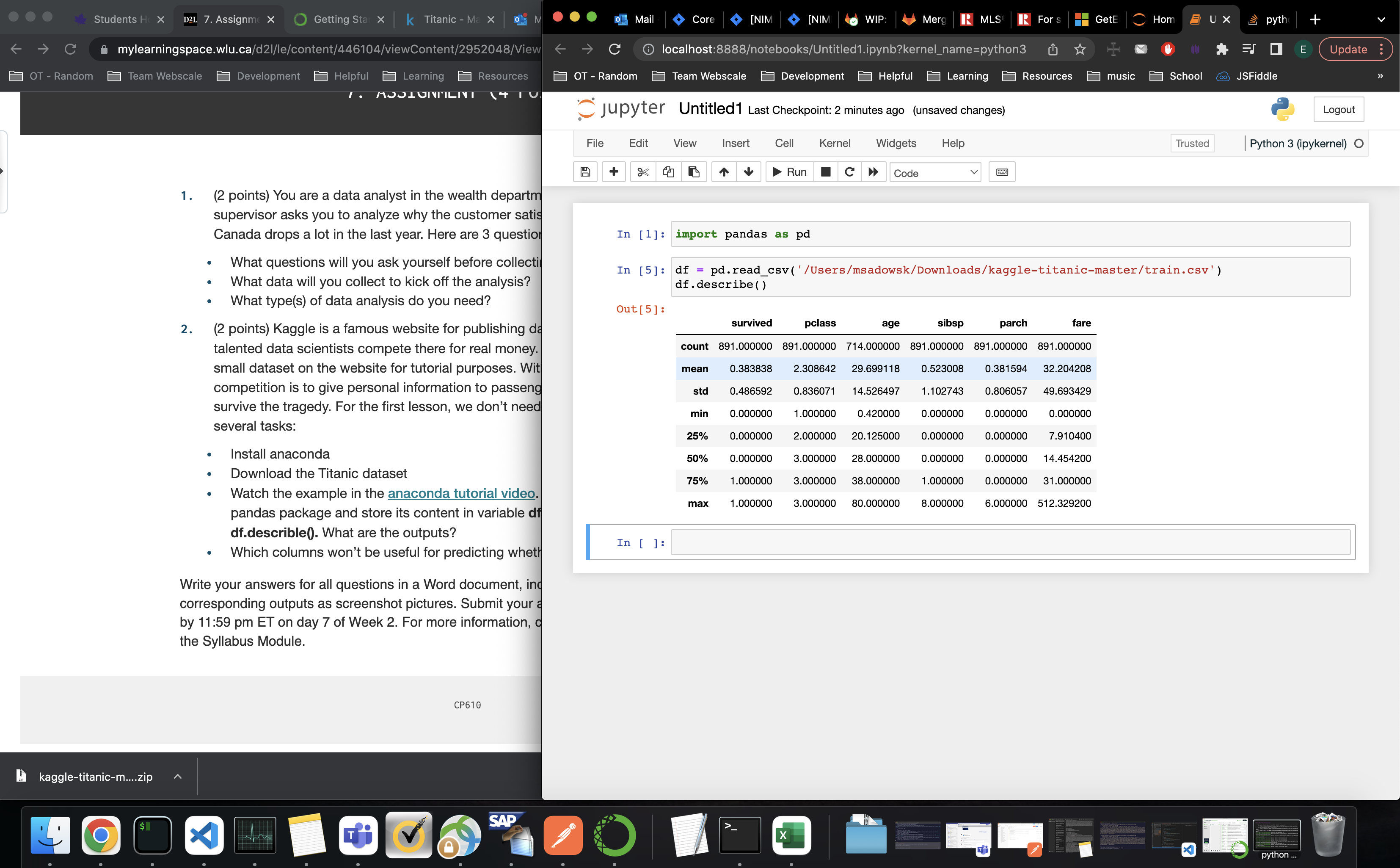
Secondly, prescriptive analysis will allow us to determine the best course of action to take to ensure this cycle of negative feedback stops. We can use the information and outcomes provided from the diagnostic analysis to determine which changes to make to solve the customer’s issues with the services.

1. **(2 points) Kaggle is a famous website for publishing data science competitions. A lot of talented data scientists compete there for real money. Titanic Data Competition is a small dataset on the website for tutorial purposes. With this dataset, the goal of this competition is to give personal information to passengers, to predict whether they can survive the tragedy. For the first lesson, we don’t need to go that far. You need to finish several tasks:**
   1. **Install anaconda**
   2. **Download the Titanic dataset**



* 1. **Watch the example in the anaconda tutorial video. Open the *train.csv* file by the pandas package and store its content in variable *df*; Try to execute *df.head(5)* and *df.describe()*. What are the outputs?**





* 1. **Which columns won’t be useful for predicting whether a passenger will survive?**

Every column provided in the data set (train.csv) has some significance to the prediction of whether the passenger will survive.

The most important columns are (in no specific order): survived (for obvious reasons), pclass (the passenger's class as it would detail who was given priority), name/sex/age (for obvious reasons), sibsp/parch (how many family members were on board, someone with no family to worry about may have a better chance at survival, etc.), and cabin (if the cabin were in the damaged part of the boat or sunk first, etc.).

The least important columns are ticket numbers/fare as these 2 do not provide any additional information we do not already have from the other columns.