**MKT 568 -Assignment 1**

**Names: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ and \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**Due: Sep 18, 11:59 PM**

**Scoring: 30 points total (15% of your grade)**

**The goal of this assignment is to help you think through what has been discussed in the course so far, apply them to real-world examples, and improve your critical thinking skills. With your team, discuss the following questions and answer them. In your submission, delete the question texts and *only* include your answers to the questions. Your submission must be written in an M.S. Word file with a Times New Roman, 12 Font. Any sources used in the assignment must be properly cited in the APA format (7th edition). Failing to cite your sources is considered plagiarism.**

**What you submit on the deadline includes (1) a Word document where you have responded to the questions listed in this document and (2) a Python code with the complete code that helps you respond to question C-2. The Python code must be in the format of a Jupyter Notebook (.ipynb).**

**This is a group project, so only one student from each team should submit these documents. After one of you submits it, the rest of the team should automatically see that they’ve submitted the assignment. If this doesn’t happen, shoot me an email.**

1. In order to understand more about customer satisfaction, the Macy’s company asked 3000 of their customers across the U.S. to fill out a questionnaire. The questionnaire included questions measuring customers’ attitudes towards Macy’s, their satisfaction with Macy’s products, customer service, and locations, as well as demographic variables. Read the following variable definitions and determine the appropriate scale for each variable. On what scales were these variables measured (Nominal, Ordinal, Interval or Ratio)? (1 pt each)
   1. Variable name: Repurchase\_Intentions

Consumers’ intentions to repurchase from Macy’s was measured by asking them this question:

How likely would you be to make another purchase from Macy’s?

1 (Not at all) 2 3 4 5 6 7 8 9 10 (Will definitely purchase again)

* 1. Variable name: Gender:

Consumers’ gender was measured by asking participants their gender with this question: What is your gender? 1) female, 2) male or 3) non-binary.

* 1. Variable name: income

Consumer’s income level was measured by asking them to write their annual household income in a text box.

* 1. Variable name: Satisfaction\_Product

Consumers were asked to indicate how satisfied they were with Macy’s product assortment?

1=Very dissatisfied, 2= moderately dissatisfied, 3=neither satisfied nor dissatisfied, 4= moderately satisfied, 5=completely satisfied

* 1. Variable name: Macys\_Location

Consumers were asked to write the name of the Macy’s location (e.g., Auburn Mall) where they usually shop.

An example of the data

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Customer ID | Repurchase\_Intentions | Gender | Income | Satisfaction\_Product | Macys\_Location |
| 1 | 5 | 3 | 100000 | 3 | Auburn mall |
| 2 | 10 | 2 | 250000 | 1 | Natick, MA |
| 3 | 3 | 1 | 32000 | 2 | Burlington Mall |
| … |  |  |  |  |  |

1. Read the following article from the Wall Street Journal about Amazon’s anticipatory shipping. Imagine you are the lead data scientist at Amazon, and you are responsible for running an analysis to anticipate shipping possibilities for Amazon consumers. (4 pts)
   1. Describe the business problem.
   2. Explain what type of data (e.g., collected from whom) and what variables (input and output) you would need to run this analysis successfully.

Article link (the article’s PDF is also uploaded to Canvas): <https://www.wsj.com/articles/BL-DGB-32082>

1. Airbnb is an online platform that connects people who want to rent out their properties with travelers looking for accommodations. It offers a range of lodging options, from rooms to entire homes, and includes user reviews to build trust. Inside Airbnb (insideairbnb.com) is a mission-driven project that provides data and advocacy regarding Airbnb’s impact on residential communities. They scrape data from Airbnb’s website and share it with the public (data scraping is the process of extracting information from websites or other sources, usually in an automated way, often using scripts or tools that gather and organize data for analysis or use in other applications.).

This dataset includes information about rental properties currently listed on Airbnb, along with their ratings, prices, and types of rental properties. This dataset can be used in various marketing strategies. One useful application is for companies that rent their properties on Airbnb. Many real estate businesses own multiple rental properties and rent them through the Airbnb platform. These businesses can use Airbnb data to calculate optimal pricing for their properties based on current listings.

I have downloaded and summarized data from Airbnb rental properties in the Boston area from the InsideAirbnb website (BostonListings.xlsx). Here is the data dictionary (definition of variables) for the BostonListings dataset

* ID: an identifier or a unique value used to distinguish individual listings in the BostonListings dataset shared with you.
* Host\_id: The Airbnb host’s id on Airbnb (what they use to log in to the system).
* Host\_name: The Airbnb host’s name as it appears on Airbnb
* Listing\_Title: The Airbnb property listing’s title.
* Neighborhood: the neighborhood in which the property is located.
* Room\_type: The type of the listed property (values include Entire home/apt, Private room, Shared room, Hotel room).
* Price: The nightly price of renting the property in U.S. dollars
* Minimum\_nights: the minimum nights that a customer can rent this property. Every time a customer moves out of a property, the owner must clean that place. Since cleaning can be costly, some property owners set a minimum night limit to ensure that customers stay at least a number of nights in the property to avoid cleaning it too much.
* Number\_of\_reviews: the number of reviews the property has received on Airbnb
* Reviews\_per\_month: the number of reviews the property has received per month since the creating of the listing on Airbnb
* Availability\_365: how many days in a year is the property available for renting. Some owners do not rent their property every day during a year for various reasons.

Please download the file, and examine it to familiarize yourself with the data and respond to the following questions based on this data.

1. For the variable in the following table determine whether this variable is a continuous variable or a categorical one. (7 pts).

|  |  |
| --- | --- |
| Variable | Categorical or Continuous? |
| Neighbourhood |  |
| Room\_type |  |
| Price |  |
| minimum\_nights |  |
| number\_of\_reviews |  |
| reviews\_per\_month |  |
| availability\_365 |  |

2. Your task in the second part of question C is to clean the BostonListings data and prepare it for analysis. We will run some analysis on this data in future sessions. What steps would you take to clean this data?

To answer this question, you must

* In the Word file, respond to this question (what steps would you take to clean the BostonListings data set). Be as detailed as possible, explain the steps you took to clean the data, and argue why you took those steps. (7 pts)
* Additionally, you must submit the Python code you wrote to clean the data when you submit the assignment. (7 pts)

Notes:

* Follow the steps of the sample Python code uploaded to Canvas (DataCleaning.ipynb) to write the required code and clean this dataset. The sample Python code is the one we worked on during the class. I added some instructions for clarity and more guidance. I recommend following this Python file to clean the data. However, if you’re pretty hands-on with data analysis, you can diverge from these instructions and clean the data your way. I will evaluate your data-cleaning approach by assessing how logical and appropriate the steps you took to clean the data were.
* When you clean the data only focus on the variables in the table included in Question C-1 (the table above). You don’t need to deal with missing values, outliers, weird values, etc. in the other columns included in the dataset.
* After completing your code and before the submission, run all code once to ensure there are no errors in the code. You will lose points if you submit a code that throws an error. To do so, in Jupyter, from the above menu, first click on **Edit-> Clear outputs of all cells** to clear all outputs. Then clock on **Run-> Run all cells** from the top menu to run all lines of code. Scroll through your code and make sure there are no errors before submission. Errors are shown in a red box below the cell that caused the error. If you see an error box, make sure you correct the code in that cell, then clear output and run all code again to make sure your code is flawless before submission.