**BIA 3713: Introduction to Business Intelligence**

**Fall 2021**

**ASSIGNMENT 1**

**Problem 1**

You have been recently hired as a Data Scientist for the urgent care clinic and will spend the next two weeks reviewing all the patient records in this dataset. There is an ID number assigned to each patient, followed by patient age and gender. After this visit, the patients were mailed a survey to gather satisfaction scores. The results are represented in the opinion column. A score of 1 is very poor, 2 is poor, 3 is neutral, 4 is good, and 5 is very good. The clinic administrator has supplied the total time each patient spent in the clinic for this visit and also the number of prior visits to the clinic for each patient. The billing department has provided the payment type and the charges billed for the visit in charges column. The patient’s admission blood pressure data come from the EHR.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **ID** | **Age (years)** | **Gender (M/F)** | **Opinion** | **Visit Time (Min)** | **Charges ($)** | **Payment Type** | **Prior Visits** | **Admission Blood Pressure** |
| **Systolic**  | **Diastolic**  |
| 1 | 33.5 | M | 5 | 64.2 | 158 | Self-Pay | 4 | 136 | 71 |
| 2 | 21.2 | F | 2 | 69.4 | 159 | Medicaid | 3 | 112 | 65 |
| 3 | 56.4 | F | 1 | 81.1 | 178 | Medicaid | 0 | 156 | 88 |
| 4 | 53.9 | M | 3 | 31.6 | 124 | Blue Cross | 1 | 125 | 80 |
| 5 | 51.2 | F | 5 | 48.5 | 146 | Aetna | 8 | 133 | 62 |

The clinic administrator will ask you questions about these data and your answers will be used to set some performance goals and strategic objectives for the clinic. Considering the small portion of the dataset provided above, answer the following questions:

1. What type of data is found in each column (e.g., categorical [nominal]; interval [continuous], etc.)?

**Age (years)-**Interval Continuous

**Gender (M/F)-**Categorical Nominal

**Opinion-**Categorical Ordinal

**Visit Time (Min)-**Interval Continuous

**Charges ($)-**Interval Continuous

**Payment Type-**Categorical Nominal

**Prior Visits-**Interval Discrete

**Blood Pressure-**Systolic Interval Continuous

**Blood Pressure-**Diastolic Interval Continuous

1. From the following numerical summary measures, which measures would you prefer to use to summarize the data in each column except for column ID?
* Mean
* Median
* Count

**Age (years) –** Mean and Median

**Gender (M/F)-**Count

**Opinion-**Count

**Visit Time (Min)-**Mean and Median

**Charges ($)-**Mean and Median

**Payment Type-**Count

**Prior Visits-**Mean and Median

**Blood Pressure-** Mean and Median

**Blood Pressure-** Mean and Median

1. From the following graphical methods, what type of methods would you prefer to use to represent the data in each column except for column ID?
* Bar Chart
* Histogram

**Age (years) –** Histogram

**Gender (M/F)-** Bar Chart

**Opinion-** Bar Chart

**Visit Time (Min)-** Histogram

**Charges ($)-**Histogram

**Payment Type-**Bar Chart

**Prior Visits-** Histogram

**Blood Pressure-** Histogram

**Blood Pressure-** Histogram

**Problem 2**

**Multiple Choice Questions**

1. A company has developed a new computer sound card whose average lifetime is unknown. In order to estimate this average, 200 sound cards are randomly selected from a large production line and tested; their average lifetime is found to be 5 years. The 200 sound cards represent a: (1 Point)
2. parameter
3. statistic
4. **sample**
5. population
6. Using data-driven insights, companies like Quartet are, surprisingly, finding that there is no relationship between improving patients’ mental and physical health together in order to improve patients’ overall health and reduce costs for the patients. (1 Point)
7. True **b) False**

1. The process of using sample statistics to draw conclusions about population parameters is called: (1 Point)
2. finding the significance level.
3. calculating descriptive statistics.
4. **doing inferential statistics.**
5. calculating the confidence level.
6. When mean gets affected by extreme outliers \_\_\_\_\_\_\_\_\_\_ can be used for finding central location: (1 Point)
7. Mode
8. **Median**
9. Range
10. None of these
11. The histogram shown below is: (1 Point)



1. Symmetrical
2. Left Skewed
3. **Right Skewed**

**Problem 3**

**Role of Data Scientist in Industry (**Refer to the article: [Data Scientist: The Sexiest Job of the 21st Century)](https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century)

1. What role do Data Scientists typically play in an organization? Discuss in about 150 words. (3.5 Points)

Data scientists are skilled professionals that use their talents and curiosity to make discoveries in the world of bid data. Data scientists typically dive into the raw data to get insights about the data. Data scientists work with large amounts of unstructured data, bring it all together, clean it, and develop visuals for a more in-depth analysis. They create models to answer critical business questions and improve their company business needs.

1. How is the role of a Data Scientist different from the role of a traditional Quantitative Analyst? Discuss with examples in about 150 words. (3.5 Points)

The role of a data scientist is different from the quantitative analyst in several aspects: quantitative analysts typically deal with structured data; data scientists, on the other hand, are the ones that work with larger amounts of unstructured data. Quantitative analysts may not need social skills - they simply analyze the data; data scientists swim into the raw data to discover insights using several tools and communicate the findings to company stakeholders in an understandable language.

1. Describe how LinkedIn, through data science, increased users’ page views and prompted them to visit more pages on its site. Discuss in about 100 words. (3 Points)

LinkedIn increased user activity and page views through various data science models. LinkedIn data scientists created models that presented users with names of people who shared common traits and interests based on their profiles. Data scientist presented users with names of people were potential connections. That model was very successful, leading to high rates of click-through and user activities.

**Problem 4**

Refer to the article: [Seven Use Cases for Data Science](https://medium.com/coriers/7-use-cases-for-data-science-and-predictive-analytics-e3616e9331f9) and Predictive Analytics

Some companies, such as Buxtonco, use data science to help predict the best retail location for a store so customers can achieve greater success and growth by understanding their customer. In general, what factors do these companies use to determine the best location? Discuss your answer in about 100 words. (3 Points)

To determine the best location of a specific store, companies look for "where" and “when” customers could spend their time and what activities are done at these places. Places that have large or significant crowds can be a good sign for the store location. The more people are in a particular area, the more customer a store can have. Understand the day of the week/month and time customers spend at a place is important to determine the location of a store. Also, it is important to know what kinds of activities customers do in certain areas. A good store location should be in places where customers will spend a significant amount of time and engage in a particular activity that could trigger them to visit the store.

**Problem 5**

Refer to the article: [Disney Uses Big Data, IoT And Machine Learning To Boost Customer Experience](https://www.forbes.com/sites/bernardmarr/2017/08/24/disney-uses-big-data-iot-and-machine-learning-to-boost-customer-experience/#13eed3eb3387)

This article describes multiple projects being developed to change the behemoth that is Disney. List and briefly describe one of these projects. Discuss your answer in about 100 words. (3 Points)

Disney Projects:

1. MagicBand Wristband
2. Shoe/Sole Recognizer
3. Movie Sentiment Analysis

The MagicBand wristband uses Radio Frequency Identification (RFID) technology that communicates and streams real-time data to multiple sensors and systems. The wristband acts as hotel keys, credit cards, tickets, FastPasses, resulting in less wait time and a better guest experience. MagicBand wristband monitors consumer behaviour, analyses purchasing habits, and provides Disney employees with real-time data.