

Master of Science (MS) in Data Science

Module: ITC6004A1 – Data Visualization

Instructor: Konstantinos Bloutsos

Student’s name: Alkiviadis Kariotis

ACG Identification Number: 241735

Term: Winter Term 2023

Type: 1st Assignment

Submission Date: Friday, January 27, 6:30 PM

**Data Visualization**

**Assignment #1.1**

You have recently joined company XYZ as a Data Scientist to staff a new team of 5 persons under the IT department which is responsible for the Data Analytics and Reporting in the organization.

Last November, the IT Director has requested some budget for 2023 to expand the team with more people and/or technology. The CIO has asked some justification for this and since you are the most experienced person in the team, the IT Director has asked from you to prepare a short summary of the value of Data Analytics could bring in the organization.

The summary should not exceed the area below (around 200-300 words).

**Answer:**

As we are experiencing the fourth industrial revolution, digitalization and the internet of things make everything interconnected which results in the production of bigger amounts of data each year in comparison with the previous ones. So, after an organization follows the entire process of acquiring, storing, and modeling the data, then we reach the point of data analysis and reporting.

Data analytics can provide valuable insights to an organization, bringing significant value to areas such as customer behavior, market trends and operational efficiency. By analyzing the aforementioned large datasets, the department can aid towards the identification of patterns and trends that could be useful in decision making at various levels of the organization. For example, the reporting dashboards that the team can provide can either be operational, strategic, or analytical aiming to serve different purposes and aid different processes throughout the company’s infrastructure.

Additionally, through data analytics overall effectiveness of the operations could be enhanced, spotting out areas for cost savings and maybe new possible revenue streams. Furthermore, it can help our company XYZ to gain a deeper understanding of our customer’s needs and wants, leading at the improvement of their engagement and satisfaction through developing more targeted and effective marketing campaigns but also differentiating XYZ from its competitors. Another important field data analytics can touch upon, has to do with risk management and compliance, meaning that the department can aid towards identifying potential risks, frauds, security threats and then the developing of strategies as to mitigate them.

Investing and expanding the team of the department will make possible to increase the overall number and accuracy of the reports delivered to any department needed, improving their efficiency and effectiveness in the decision-making process as well as aiding towards insights identification by eliminating data silos. This will enable to harness the full potential of data analytics and drive the success and growth of the organization.

**Data Visualization**

**Assignment #1.2**

The IT Director who is leading the Data Analytics team believes that we can expand the team in 2023 with more people. In that sense and because he would like to train the new joiners quickly and efficiently, he has asked from you to prepare a draft document with guidelines on Data Quality Assessment.

The purpose of this document is to specify what a junior Data Analytics should do in order to assess the quality of the data that he/she is going to work with.

The draft should not exceed the 2 pages (around 400-500 words).

**Answer:**

The starting point of the job of a Data Analyst is to ensure the quality of the data before starting any analysis or modeling. That’s because if the data that we possess are low quality being noisy, incomplete, or inconsistent for example, then inevitably the results will be based on faulty predictions leading to incorrect conclusions and ultimately wrong and uninformed decisions about the organization.

The purpose of the following guidelines is to provide to new employees the steps that should be taken in order to assess the quality of the data that they have acquired, resulting in a more efficient and effective daily basis, minimizing the need to go back and forth the same procedures and processes. But before taking into consideration all the step procedure, the Junior Data Analyst should first focus on understanding the data that he/she has acquired. Meaning to gain a “deeper” knowledge on the business context that they are going to be used for, understand their structure, the relationship between the tables and finally the process that was used to collect those data.

So, the steps are:

1. Completeness: check if the data meet the expectations, meaning whether the provided data contain all the required columns with no large number of missing values in them, even though the optional data could be incomplete.
2. Uniqueness: check whether data have duplicate values, meaning if there are rows that provide information on the same variable or entity.
3. Consistency: check whether data from similar data sources follow the same structure and reflect the same information. Meaning whether there are inconsistencies between related tables and unexpected values and outliers.
4. Conformity: check whether the values of the needed and maybe optional columns of the dataset are inserted in the correct format, size, and data type.
5. Accuracy: check whether the data actually reflects real world objects or events described by comparing it with external sources or known facts. Also, in this step we can run the data through some pre-defined rules and additionally make some basic calculation or cross checking with older similar datasets.
6. Documentation: record steps that were performed towards the assessment process, mentioning any potential issues that were found and the methods used to resolve them, and finally the ways that were used to evaluate the data.

By sticking to these guidelines, the Junior Data Analyst will gain a better idea whether the data is accurate, complete, and ready for analysis. This will end up on informed decisions based on the data, and the organization to trust the insights generated by the department. Finally, it is important to state that data quality assessment is an ongoing process which should be reviewed regularly just to make sure data remains accurate and relevant.