Georgia State University

CSC4780/6780&DSCI4780 – Fundamentals of Data Science

*[Spring 2024]*

Project Progress Report

[H2O Analytics]

[H2O Analytics]

[Sebastien Lee]

[Emery Hagye]

[Julian Ornelas]

**Table of Contents**

[**1 Business Understanding**](#_heading=h.gjdgxs) **3**

[1.1 Business Problem](#_heading=h.bnkyogcs40q6) 3

[1.2 Dataset](#_heading=h.hwr14ta6o0cx) 3

[1.3 Proposed Analytics Solution](#_heading=h.iy5juclrub6j) 4

[**2 Data Exploration and Preprocessing**](#_heading=h.913hk1g9ex9q) **5**

[2.1 Data Quality Report](#_heading=h.vrbdicq01d9p) 5

[2.2 Missing Values and Outliers](#_heading=h.a9e5lucq94ha) 6

[2.3 Normalization](#_heading=h.4f3b3tkc9ksx) 6

[2.4 Transformations](#_heading=h.iy31tdg94rhj) 7

# 1 Business Understanding

*[REMOVE THIS – Briefly describe your project topic, business problem, what is your motivation for selecting this project, how your dataset can be used, and some of the important aspects of your analytics design. You are required to identify your target variable (or overall goal of your analytics solution) and talk about its characteristics. It should include Business Problem, Dataset, and Analytics Solution subsections.]*

## 1.1 Business Problem

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Non sodales neque sodales ut etiam sit amet. Quis commodo odio aenean sed adipiscing diam donec adipiscing. Nulla porttitor massa id neque. Auctor eu augue ut lectus arcu bibendum at. Consectetur lorem donec massa sapien faucibus et. Mauris vitae ultricies leo integer malesuada nunc. Feugiat nibh sed pulvinar proin gravida hendrerit lectus. Feugiat in ante metus dictum at tempor commodo. Nec feugiat nisl pretium fusce id. Mollis nunc sed id semper risus in hendrerit. Risus quis varius quam quisque id diam vel quam elementum. Vitae suscipit tellus mauris a. Iaculis urna id volutpat lacus laoreet.

Nunc lobortis mattis aliquam faucibus purus in massa tempor. Ultrices gravida dictum fusce ut. In mollis nunc sed id semper risus. Eget mi proin sed libero enim. Et tortor consequat id porta. Vitae nunc sed velit dignissim. Tortor at auctor urna nunc. Blandit massa enim nec dui nunc mattis enim ut tellus. In est ante in nibh. Nunc eget lorem dolor sed viverra ipsum nunc. Commodo viverra maecenas accumsan lacus vel. Purus semper eget duis at. Augue mauris augue neque gravida in fermentum et. Sagittis purus sit amet volutpat consequat mauris nunc. Commodo quis imperdiet massa tincidunt nunc pulvinar sapien et ligula. Mauris sit amet massa vitae tortor condimentum. Cras tincidunt lobortis feugiat vivamus at augue. Vitae et leo duis ut diam quam nulla. Rutrum tellus pellentesque eu tincidunt tortor. Turpis egestas maecenas pharetra convallis posuere.

## 1.2 Dataset

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Non sodales neque sodales ut etiam sit amet. Quis commodo odio aenean sed adipiscing diam donec adipiscing. Nulla porttitor massa id neque. Auctor eu augue ut lectus arcu bibendum at. Consectetur lorem donec massa sapien faucibus et. Mauris vitae ultricies leo integer malesuada nunc. Feugiat nibh sed pulvinar proin gravida hendrerit lectus. Feugiat in ante metus dictum at tempor commodo. Nec feugiat nisl pretium fusce id. Mollis nunc sed id semper risus in hendrerit. Risus quis varius quam quisque id diam vel quam elementum. Vitae suscipit tellus mauris a. Iaculis urna id volutpat lacus laoreet.

Nunc lobortis mattis aliquam faucibus purus in massa tempor. Ultrices gravida dictum fusce ut. In mollis nunc sed id semper risus. Eget mi proin sed libero enim. Et tortor consequat id porta. Vitae nunc sed velit dignissim. Tortor at auctor urna nunc. Blandit massa enim nec dui nunc mattis enim ut tellus. In est ante in nibh. Nunc eget lorem dolor sed viverra ipsum nunc. Commodo viverra maecenas accumsan lacus vel. Purus semper eget duis at. Augue mauris augue neque gravida in fermentum et. Sagittis purus sit amet volutpat consequat mauris nunc. Commodo quis imperdiet massa tincidunt nunc pulvinar sapien et ligula. Mauris sit amet massa vitae tortor condimentum. Cras tincidunt lobortis feugiat vivamus at augue. Vitae et leo duis ut diam quam nulla. Rutrum tellus pellentesque eu tincidunt tortor. Turpis egestas maecenas pharetra convallis posuere.

## 1.3 Proposed Analytics Solution

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Non sodales neque sodales ut etiam sit amet. Quis commodo odio aenean sed adipiscing diam donec adipiscing. Nulla porttitor massa id neque. Auctor eu augue ut lectus arcu bibendum at. Consectetur lorem donec massa sapien faucibus et. Mauris vitae ultricies leo integer malesuada nunc. Feugiat nibh sed pulvinar proin gravida hendrerit lectus. Feugiat in ante metus dictum at tempor commodo. Nec feugiat nisl pretium fusce id. Mollis nunc sed id semper risus in hendrerit. Risus quis varius quam quisque id diam vel quam elementum. Vitae suscipit tellus mauris a. Iaculis urna id volutpat lacus laoreet.

Nunc lobortis mattis aliquam faucibus purus in massa tempor. Ultrices gravida dictum fusce ut. In mollis nunc sed id semper risus. Eget mi proin sed libero enim. Et tortor consequat id porta. Vitae nunc sed velit dignissim. Tortor at auctor urna nunc. Blandit massa enim nec dui nunc mattis enim ut tellus. In est ante in nibh. Nunc eget lorem dolor sed viverra ipsum nunc. Commodo viverra maecenas accumsan lacus vel. Purus semper eget duis at. Augue mauris augue neque gravida in fermentum et. Sagittis purus sit amet volutpat consequat mauris nunc. Commodo quis imperdiet massa tincidunt nunc pulvinar sapien et ligula. Mauris sit amet massa vitae tortor condimentum. Cras tincidunt lobortis feugiat vivamus at augue. Vitae et leo duis ut diam quam nulla. Rutrum tellus pellentesque eu tincidunt tortor. Turpis egestas maecenas pharetra convallis posuere.

# 2 Data Exploration and Preprocessing

*[REMOVE THIS – The second part of your report will be related to the data understanding. Explore and explain your attributes. Provide bar charts or histograms when necessary (i.e., if you see a pattern after exploration). Next, provide a data quality report (See slides and textbook for creating a data quality report). Describe how you handle missing values and outliers. [Describe your feature selection schema and later additional transformations.]*

## 2.1 Data Quality Report

Lorem ipsum dolor sit amet, consectetur adipiscing elit, sed do eiusmod tempor incididunt ut labore et dolore magna aliqua. Vestibulum rhoncus est pellentesque elit ullamcorper. Ut lectus arcu bibendum at. Commodo elit at imperdiet dui accumsan sit. Pharetra et ultrices neque ornare aenean. Imperdiet nulla malesuada pellentesque elit eget gravida. Dignissim sodales ut eu sem integer vitae justo eget. Cras semper auctor neque vitae tempus quam pellentesque nec. Fringilla phasellus faucibus scelerisque eleifend donec pretium. Pharetra magna ac placerat vestibulum lectus mauris ultrices. Dictum sit amet justo donec enim. Egestas integer eget aliquet nibh praesent tristique. Amet mauris commodo quis imperdiet massa tincidunt. Quis imperdiet massa tincidunt nunc pulvinar sapien et ligula. Diam maecenas ultricies mi eget mauris pharetra et ultrices neque. Iaculis nunc sed augue lacus viverra vitae.

Table 1. Data Quality Report for Categorical Features

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Lorem | ipsum | datum | dolor | sit |
| Feature 1 | 1 | 3 |  |  |
| Feature 2 | 2 | 4 |  |  |
| Feature 4 | 5 | 6 |  |  |

Table 2. Data Quality Report for Continuous Features

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Lorem | ipsum | datum | dolor | sit |
| Feature 1 | 1 | 3 |  |  |
| Feature 2 | 2 | 4 |  |  |
| Feature 4 | 5 | 6 |  |  |

Figure 1. Visualizations of Categorical and Continuous Features in Dataset

***Bar plot***

***Bar plot***

***Bar plot***

***Histogram***

***Histogram***

***Histogram***

## 2.2 Missing Values and Outliers

Diam quam nulla porttitor massa id neque aliquam. Porttitor massa id neque aliquam vestibulum. Ultricies mi eget mauris pharetra et ultrices neque ornare. Sagittis orci a scelerisque purus semper eget duis at. Tempor orci eu lobortis elementum nibh tellus. Dui nunc mattis enim ut tellus elementum. Eget mauris pharetra et ultrices neque. Sed risus pretium quam vulputate. Gravida neque convallis a cras semper auctor. Odio tempor orci dapibus ultrices.

Est ullamcorper eget nulla facilisi etiam dignissim diam quis. Neque laoreet suspendisse interdum consectetur libero. Nisl rhoncus mattis rhoncus urna neque viverra. Enim eu turpis egestas pretium aenean pharetra. Molestie nunc non blandit massa. Nulla malesuada pellentesque elit eget gravida. Nibh tellus molestie nunc non blandit massa enim nec. Ipsum faucibus vitae aliquet nec ullamcorper sit amet risus. Cras pulvinar mattis nunc sed blandit. Nunc id cursus metus aliquam eleifend mi in. Mi sit amet mauris commodo quis imperdiet massa. Vulputate enim nulla aliquet porttitor lacus luctus accumsan tortor. Varius vel pharetra vel turpis nunc eget lorem. Odio eu feugiat pretium nibh ipsum consequat nisl.

## 2.3 Normalization

Arcu bibendum at varius vel pharetra vel. Nascetur ridiculus mus mauris vitae. Consectetur libero id faucibus nisl tincidunt. Consequat id porta nibh venenatis cras sed felis eget velit. Dolor sit amet consectetur adipiscing elit. Consequat semper viverra nam libero justo laoreet. Purus sit amet volutpat consequat mauris. Quis blandit turpis cursus in. Nec sagittis aliquam malesuada bibendum arcu. Pellentesque nec nam aliquam sem et tortor consequat id. Quam quisque id diam vel quam elementum pulvinar etiam. Facilisis magna etiam tempor orci eu lobortis elementum. Nec sagittis aliquam malesuada bibendum arcu vitae.

## 2.4 Transformations

Est ullamcorper eget nulla facilisi etiam dignissim diam quis. Neque laoreet suspendisse interdum consectetur libero. Nisl rhoncus mattis rhoncus urna neque viverra. Enim eu turpis egestas pretium aenean pharetra. Molestie nunc non blandit massa. Nulla malesuada pellentesque elit eget gravida. Nibh tellus molestie nunc non blandit massa enim nec. Ipsum faucibus vitae aliquet nec ullamcorper sit amet risus. Cras pulvinar mattis nunc sed blandit. Nunc id cursus metus aliquam eleifend mi in. Mi sit amet mauris commodo quis imperdiet massa. Vulputate enim nulla aliquet porttitor lacus luctus accumsan tortor. Varius vel pharetra vel turpis nunc eget lorem. Odio eu feugiat pretium nibh ipsum consequat nisl.