CS777 – Term Project Proposal Submission Template

**Student: Aaron Taylor**

1. Data set description: Provide a detailed description of the public data set you have selected, including its source, format, and any relevant details about the data.

|  |
| --- |
| *Your description should go here.*  Small Taxi dataset: <https://storage.googleapis.com/met-cs-777-data/taxi-data-sorted-small.csv.bz2>  This is the dataset that was used for assignments 1 & 3. |

1. Research question: Clearly define your research question and explain why studying is important. What do you want to learn from the data?

|  |
| --- |
| *Your description should go here.*  Where should taxi hubs be centered? Or where can a taxi go when they need a ride? This will help taxi drivers increase their rides and should increase the money they earn. |

1. Machine Learning model: Specify the type of machine learning model you plan to use, such as classification or clustering, and explain why you have chosen this model.

|  |
| --- |
| *Your description should go here.*  I will use K-means clustering to create clusters based off ride longitude and latitude. This will allow me to see which areas have the most riders. |

1. Expected outcomes: What do you expect to achieve after implementing your learning model? What do you hope to learn or discover from your data analysis?

|  |
| --- |
| *Your description should go here.*  I hope this will give insight into areas New York taxi drivers can go for a guaranteed ride. |

1. Evaluation plan: Explain how you plan to evaluate your project and assess the correctness of your model. What metrics or methods will you use to evaluate the effectiveness of your learning model? How well do you expect the model to work, and how will you measure its performance?

|  |
| --- |
| *Your description should go here.*  To evaluate the model I will use a silhouette score to quantify the quality of the model, as well as see which clusters are the largest to indicate hotspots. |