Analyzing the NYC Subway Dataset Documentation

Release 1.0

Ignacio Toledo

December 29, 2014

CHAPTER

ONE

OVERVIEW

This project consists of two parts. In Part 1 of the project, you should have completed the questions in Problem Sets 2, 3, 4, and 5 in the Introduction to Data Science course.

This document addresses part 2 of the project. Please use this document as a template and answer the following questions to explain your reasoning and conclusion behind your work in the problem sets. You will attach a document with your answers to these questions as part of your final project submission.

CHAPTER

TWO

STATISTICAL TEST

2.1 Statistical Test Used

2.2 Justify the Statistical Test

2.3 Results

2.4 Interpretation

CHAPTER

THREE

LINEAR REGRESSION

3.1 Linear regression algorithm(s)

3.1.1 Gradient descent

3.1.2 OLS (with statsmodels)

3.2 Features used

3.3 Feature selection: why?

3.4 Results: R Squared

3.5 Interpretation and limits

CHAPTER

FOUR

VISUALIZATION

4.1 Ridership distribution with weather

4.2 Supporting visualizations

CHAPTER

FIVE

CONCLUSION

CHAPTER

SIX

REFLECTION

6.1 Shortcomings and limitations

6.2 Insights