U D A C I T Y

PROJECT SPECIFICATION

NYC Subway Data Analysis

Data Gathering

CRITERIA	MEETS SPECIFICATIONS
Exercise 1.1 - Collect files from June 2017 on the New York City Subway website using Python.	4 files were collected and saved using the Python libraries responsible for data collection on the Internet.
Exercise 1.2 - Consolidate the 4 files into a single file.	The 4 files became 1 single file, preserving all columns in the source file.
Ex. 1.3 to 1.6 - Load the information into a Pandas Dataframe and format the date and time columns for the New York City subway data.	Data was loaded into a Panda Dataframe. Only rows where DESCn has the regular value were selected. The Existsn and Entriesn columns were created correctly. The time column was formatted as an integer.

Data Analysis

CRITERIA	MEETS SPECIFICATIONS
Ex. 2.1 to 2.2 - Was the number of rainy days entered correctly?	Was Python used to calculate the number of rainy days? Was the value calculated correctly?
Ex. 2.3 - Was the maximum temperature for cloudy days calculated correctly?	Was Python used to calculate the maximum temperature for cloudy days? Was the value calculated correctly?
2.4 to 2.5 - Were means calculated correctly?	Code that displays the mean responses should display the correct answers.

CRITERIA	MEETS SPECIFICATIONS
Ex. 2.6 to 2.7 - Were the histograms plotted? Were distribution findings correct?	Python was used to plot the histograms. The histograms were plotted to facilitate data distribution identification. Data distribution was identified correctly.

MapReduce

CRITERIA	MEETS SPECIFICATIONS
Ex 3.2 - The Reducer was correctly created.	The Reducer output must comply with the file requirements.
Ex 3.1 - The Mapper was created correctly.	The Mapper output must comply with the file requirements.
Ex. 3.2 - The Mapper and Reducer code runs with no bugs.	The Reducer and Mapper code must run with no bugs.

Untitled

CRITERIA	MEETS SPECIFICATIONS

Student FAQ