Jimmy Nam

10/08/2025

DATA 5100

Seattle Weather Project

Report

The project is to perform data science methodology to examine prescription data and determine whether it rains more in Seattle, WA or New York City, NY.

First, I collected the data set of New York City’s precipitation reports specifically in JFK airport station from National Oceanic and Atmospheric Administration.

I started with inspecting the data set. Checking the columns and rows to gain full insight of what each columns represents and mean. I would ask which columns are something that I need to look at or which columns are not relevant to the research I’m doing for this project.

A graph of different colored bars

AI-generated content may be incorrect.

This boxplot examines the monthly variation in precipitation between Seattle and New York City. While both cities are known for their rainfall, their patterns differ substantially in timing and frequency.

Methods:

A graph of different colored bars

AI-generated content may be incorrect.

This

Methods: Data cleaning and preparation were performed using **Python** (pandas and NumPy). Missing values and duplicates were removed, and dates were standardized. Each dataset was grouped by month and city, then summarized to calculate the proportion of wet days.

A bar chart was produced using **Seaborn**, with error bars representing one standard error of the mean. Asterisks (\*) mark months where the difference in precipitation frequency between the two cities was statistically significant (p < 0.05), based on two-sample t-tests.

This method emphasizes general patterns in rainfall frequency rather than precise volume, making the results interpretable for a general audience.

A graph of different blue squares

AI-generated content may be incorrect.

This bar graph compares the frequency of precipitation between Seattle, WA and New York City, NY using