The Price of Healthcare:
Does **a Higher GDP**
Guarantee **a Better Patient Experience**?

Kevin Tran
Mini Project
DATA400-01, Fall 2025

Table of
Contents

Introduction

- ECON three-course sequence
- Many cases of travelling back and forth between countries for treatment

→ Question:

"Where should a middle-income family choose for serious illness treatment?"

Data Range

- Original Idea: Western vs. Eastern, e.g. US vs. Asia

  - In my opinion, a bit too broad

- Focused: in **one** major US city and **one** major Asian city.

 - New York and Singapore (the capital of Singapore).

Data Retrieval

- Scraping reviews of Hospital using "General Hospital in [...]"

  ○ In this case, "New York, NY" and "Singapore, Singapore"

  ○ Was not quite stable, sometimes it fails

- Then I combined what is readily available after scrapes and export it into a .csv file so it can be processed faster with Excel.

  ○ Of course it is possible to use Python/Anaconda right away but with this amount of data I figured out it would be faster to just use Excel.
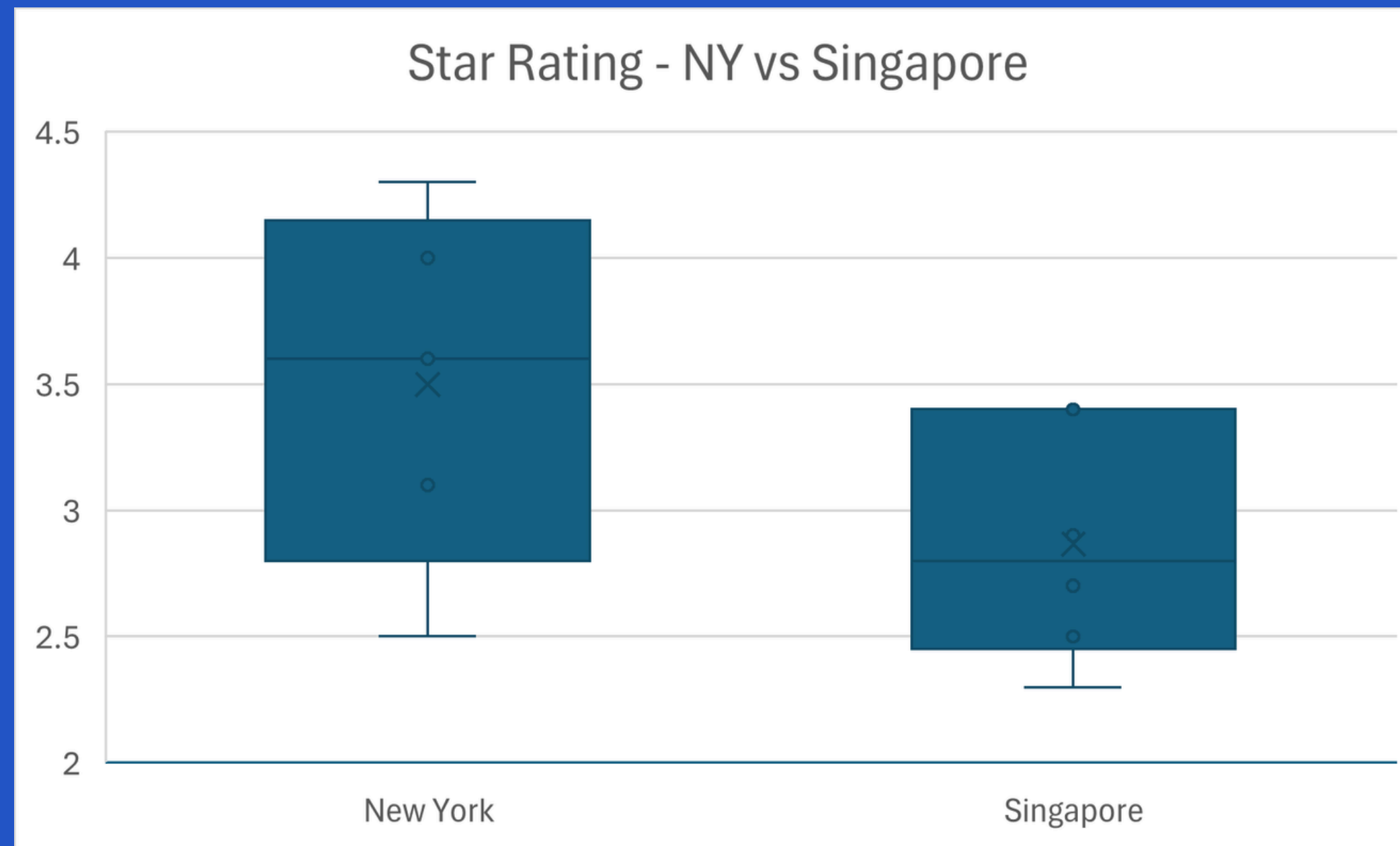
Exploratory Data Analysis (EDA)

- From my data:
  - Average star rating of NYC is around 0.6 higher than that of Singapore
  - Reviews of NYC is more diversed, while Singapore is more similar

City/Stat	Mean	Median	Std.Dev
NY	3.5	3.6	0.718
Singapore	2.867	2.8	0.459

NY vs. Singapore Statistics
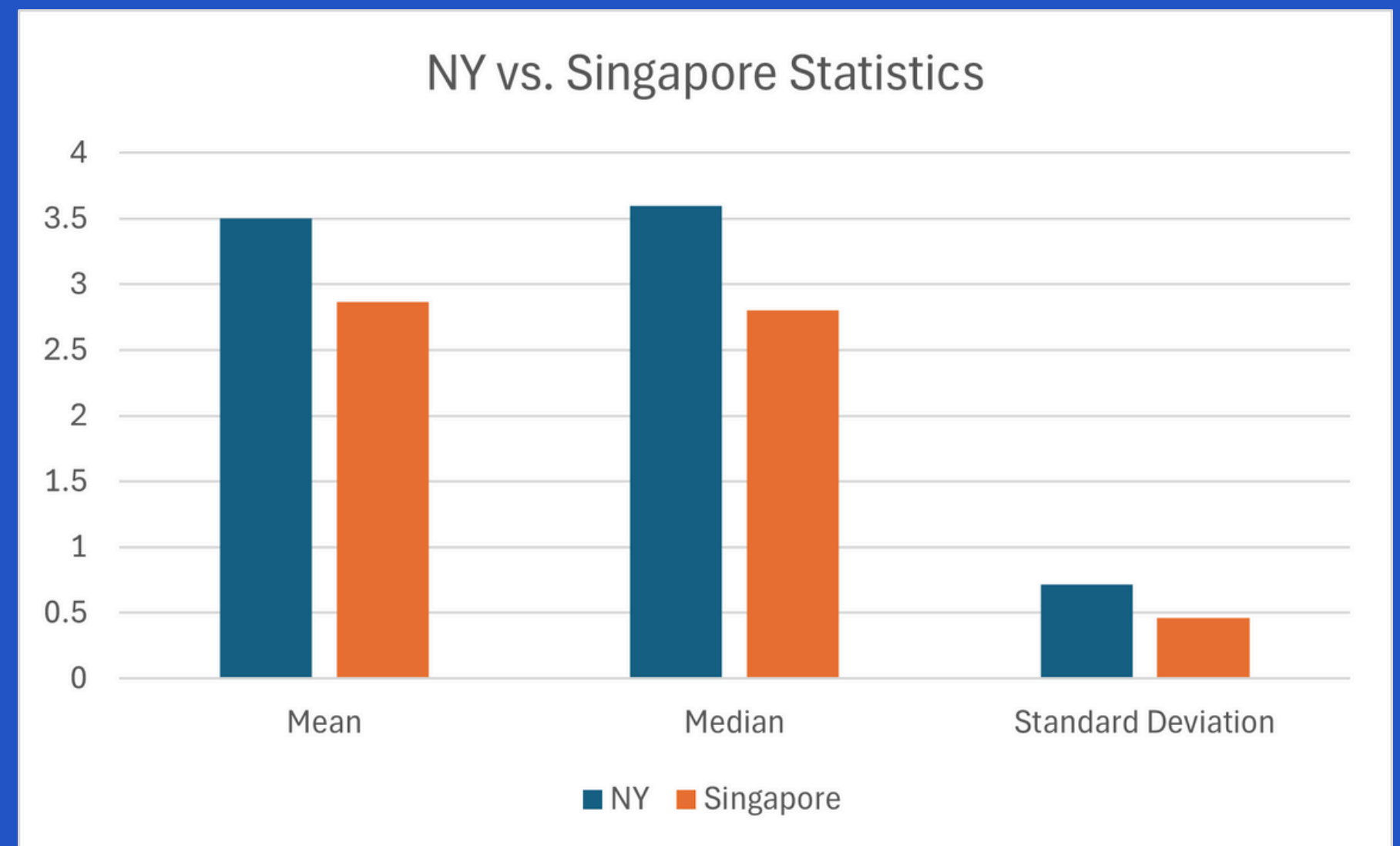
Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA)

Number of Review vs Star Rating - NY & Singapore
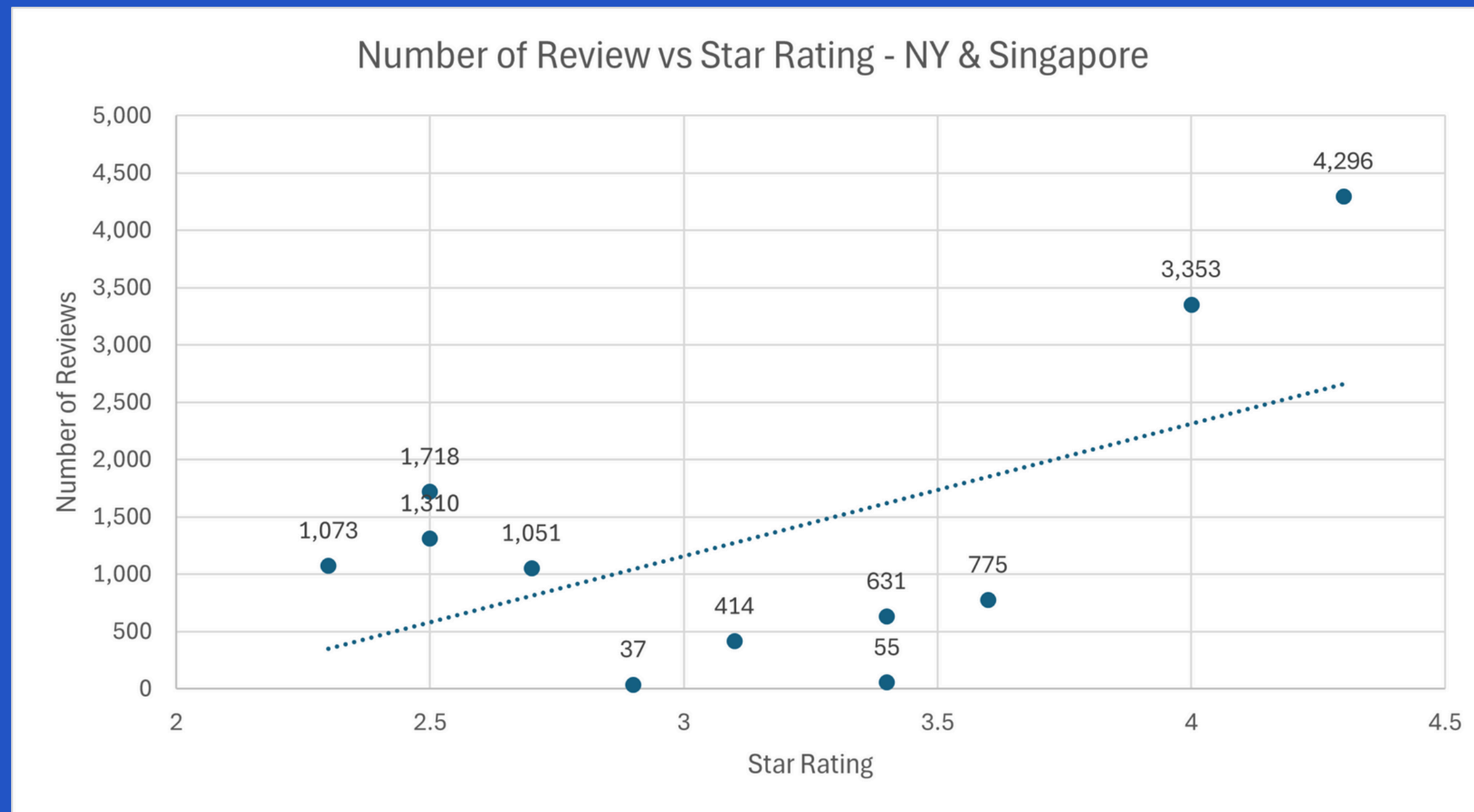
Exploratory Data Analysis (EDA)

NY General Hospital Review Word Cloud

Exploratory Data Analysis (EDA)

Singapore General Hospital Review Word Cloud

Exploratory Data Analysis (EDA)

GDP Per Capita, Nominal (2024)

- New York: $116,883

- Singapore: $90,674

→ While the wealthier city, New York, does have a higher average star rating, it also shows much greater inconsistency in patient satisfaction. Furthermore, the qualitative data suggests that patients in New York are more concerned with systemic issues like long wait times.

Exploratory Data Analysis (EDA)

Summary of Key Findings

- In this sample, New York hospitals had a higher average star rating (3.5) than Singapore hospitals (2.87), but also showed significantly more variability in patient ratings.

- The word clouds suggest that reviews for New York hospitals more frequently mention themes of wait times (e.g., "wait," "hours," "emergency"), while Singapore reviews contain more neutral, process-oriented terms (e.g., "appointment," "registration").

- While both New York and Singapore are high-GDP locations, the data suggests the city with the slightly higher economic output (New York) did not offer a uniformly better patient experience, particularly regarding efficiency.

- This preliminary analysis suggests that a higher GDP does not necessarily guarantee a better or more consistent patient experience.

Implications **for Stakeholders**

- Patients and Families
  - Informed Decision-Making
  - Risk of Misinformation
- Hospitals and Healthcare Providers
  - Performance Benchmarking
  - Reputational Impact

- Policymakers and Governments
  - Identifying Systemic Issues
  - Informing Health Policy
- Insurers and Medical Tourism Agencies
  - Network Development
  - Product and Package Design

Ethical Implications

- *Cultural* Reporting Bias: satisfaction & dissatisfaction **varies significantly between cultures.**

- *Representation* Bias: collected data will likely represent some groups of patients more. (**over-represent**)
  - Younger
  - Wealthier
  - Urban populations, with reliable internet and **cultural** habit of leaving online reviews

- *Oversimplification*: simplifying a complex issue with many unpredictable variables might promote cultural stereotypes, and unfairly criticize an entire nation's healthcare workforce.

Societal Implications

- *Misleading* Patients: The focus of the project is to help a middle-income family choose where to seek treatment. If analysis is flawed due to the biases, conclusions could **misinform** these families, potentially leading them to make poor healthcare decisions.

- *Correlation* vs. *Causation*: A relationship between a country's GDP and its healthcare quality is a **correlation**, not a **causation**. Healthcare systems are complex products of history, culture, and politics that economic data alone cannot explain. Presenting a correlation as a cause could lead to flawed policy recommendations.

- Reinforcing *Stereotypes*: An analysis that isn't handled with extreme care could reinforce existing negative stereotypes about developing countries, contributing to a biased international perception of their capabilities and quality of life.

Legal Implications

- Terms of Service (ToS): depends on each website, some might not allow scraping and violating their terms might have legal consequences.
- Data Privacy: Although online reviews are public, they can sometimes contain sensitive personal or health information.
  - Therefore, proper procedure must be performed to ensure the data is anonymized and no personally identifiable information is stored or published.

Future Expansions (maybeee...)

Gathering review posts from **Reddit, Quora and more** to perform further analysis

Expand data range to cover larger areas

Implement an NLP Model & Sentiment Analysis

Proceed with a multivariate regression model, with the output from NLP Model added.

Thank You

Questions &/ Suggestions?