Project Requirements:
https://dsan.georgetown.domains/dsan5100/specific/project.html

Research Question: What are the key factors contributing to high housing risk scores across districts, and how do public health and socioeconomic indicators correlate with these risks?

Datasets:
https://datawrapper.dwcdn.net/XDEVF/11/

Features:

- Covid vax & death rates
- Uninsured percent
- Crowding percent
- Area median income percent
- POC percent
- Rent burden percent
- Eviction filings rate
- Tenant cases rate
- Housing violations rate
- Unplanned outages rate
- Change in median gross rent
- Change in sale price
- Number of new unaffordable units
- Rate for foreclosure filings
- Nonbank share small home loans
- Etc.

Tasks

Clean/EDA	Amanda
Public Health Bureau Covid Vax Covid Death Rates Uninsured Percent	Alivia

Socioeconomic Area median income percent POC percent Change in median gross rent Change in sale price Number of new unaffordable units Nonbank share small home loans	Satomi
Housing (eviction, rent burden, foreclosure filing, etc) Crowding percent Rent burden percent Eviction filings rate Tenant cases rate Housing violations rate Unplanned outages rate Nonbank share small home loans	Jeffrey
High Housing Risk, look at categorical data maybe, borough analysis	Mandy

** We might need additional datasets since this is only <100 rows and aggregated data. Going to try to go through the database provide through Github and extract complaints

Statistical Methods:

- Bayesian
    - Estimate probabilities of high-risk outcomes given specific conditions
- Monte Carlo Simulation
    -
- Hypothesis Testing
    - Comparing mean risk scores across boroughs using t-tests or ANOVA or both!
    - Categorical values use chi square tests
    - Using permutation tests or bootstrap or both to validate findings

- MLE
    - Estimate relationships between variables ?
- Expected Value
    - Expected risk score under diff conditions

Presentation Breakdown

- Intro
- EDA - Talk about general findings
- Intro+ EDA would 3 mins
- Public Health
- Statistical Analysis 3
- Talk the about implications of the results
- Takeaway statements
- Conclusion
- Reference

Timeline:

- Dec 1 - Meet and delegate tasks
- Dec 2 - Analysis Finished (BY MONDAY NIGHT) Zoom @10pm
- Dec 3 - Presentation created (BY TUESDAY NIGHT)
- Dec 4 - Before class prepare for the presentation @1 pm @DSAN Lounge
- Dec 4 - Present
- Dec 11 - Final Report Due

Timeline:

November 24, 2024 - Clean datasets, prelim EDA

- Talk about significant data points, figure out next steps, split hypothesis testing work

November 27, 2024 - EDA / hypothesis testing & interpret results

December 2, 2024 - Finish Presentation / practice

December 4, 2024 - Presentation due

December 5, 2024 - Delegate sections, work on report

December 11, 2024 - Final Report due