CIS 9655

Project 2
Let Me Explore a Hypothesis

Objective:

The goal of this project is to develop and explore a short, data-informed hypothesis in a specific area (such as healthcare, finance, climate change, social media, education, public policy, or sports analytics).

- Gather and explore relevant datasets.
- Create effective, interactive visualizations using Tableau to allow the reader to engage with your findings.
- Focus on exploration, not simply proving or disproving your hypothesis. As opposed to Project 1 (which was an explanatory report), this project invites the audience to draw their own insights through interaction with your visualizations.

Proposal Requirements:

Each team or individual must submit a written project proposal (200–400 words) addressing the following:

- **Research Question or Hypothesis:** What will you investigate using Tableau? (Frame this clearly, even if exploratory.)
- **Motivation and Relevance:** Why is this hypothesis interesting, meaningful, or timely?
- **Data Sources:** What dataset(s) will you use, and why are they appropriate? Include direct links to the dataset(s).

Dataset Requirements:

Data Sources:

You must use open, public datasets — No Kaggle or pre-curated competition datasets. Look for local (NYC), state (NY), federal (US), or international (UN, World Bank, etc.) sources.

Your dataset must meet these minimum requirements:

Requirement	**Details**
Wide Dataset	At least 10 unique variables. ● 2–3 must be quantitative (numerical) ● 1–2 must be temporal (time-based) ● 1–2 must be geographic (zip code, lat/long, neighborhood, etc.). ● The remaining variables may be categorical or ordinal.
Long Dataset:	At least 10,000 observations (records/rows).
Multiple Datasets (if needed):	You may use multiple datasets and join them if needed.
Geospatial Data:	Include a shapefile or geoJSON related to your geographic fields. *Tip:* You may need to join the geospatial file to your dataset using a shared geographic key (e.g., ZIP code).

Expectations for Tableau Visualizations:

Your visualizations must allow interactive exploration — not just static displays. Incorporate elements like:

- Filters
- Parameter controls
- Highlight actions
- Hover tooltips
- Dashboards combining multiple views

Strive to encourage reader-driven discovery: users should be able to slice and explore different patterns on their own.

Notes and Reminders:

- Hypotheses can be exploratory: you are allowed to pose open-ended questions.
- Dataset preparation is part of your grade: cleaning, transforming, or merging data into a usable form is expected. Document your data preparation choices briefly.
- Document any use of genAI.

Grading Rubric

Criteria	Excellent	Satisfactory	Needs Improvement
Hypothesis Quality (5 points)	*5 points* Hypothesis is clear, thoughtful, exploratory, and relevant; invites meaningful exploration.	*3 points* Hypothesis is stated but somewhat vague or conventional; exploration is limited.	*1 point* Hypothesis is unclear, trivial, or inappropriate for exploratory analysis.
Dataset Suitability and Preparation (25 points)	*25 points* Dataset(s) fully meet size, variable, and structure requirements; thoughtful data preparation and cleaning are evident and documented.	*18 points* Dataset(s) mostly meet requirements with minor gaps; basic preparation is done but documentation is limited.	*10 points* Dataset(s) are insufficient (too small, too narrow, or missing key fields); little or no evidence of data preparation.
Visualizations (Interactivity and Design) (40 points)	*40 points* Visualizations are highly interactive (filters, tooltips, parameters) and designed thoughtfully to encourage user exploration.	*30 points* Visualizations have basic interactivity but lack depth or design polish; some features may be incomplete.	*15 points* Visualizations are static, difficult to navigate, or fail to allow meaningful exploration.
Analysis and Exploration Depth (20 points)	*20 points* Visualizations uncover multiple layers of insights; users can clearly engage with patterns and relationships through interaction.	*15 points* Some insights available through interaction, but patterns are shallow or inconsistently explored.	*8 points* Little or no new insight generated; visualizations do not meaningfully support exploration.
Communication and Presentation (10 points)	*10 points* Clear, professional final dashboard or workbook; logical flow and layout; appropriate use of annotations and guides for users.	*7 points* Mostly clear presentation; some layout or communication issues but overall understandable.	*4 points* Disorganized or confusing dashboard; lacks sufficient guidance for reader exploration.

Sample Proposal

Research Question:

How has the distribution of 311 noise complaints in NYC changed over time, and are there significant patterns by borough?

Motivation and Relevance:

Urban noise is a major quality-of-life issue, especially in a dense city like New York. Understanding trends and regional differences in noise complaints can help inform public policy and resource allocation for city services.

Data Sources:

Primary dataset: NYC Open Data - "311 Service Requests from 2010 to Present" (over 20 million records)
Secondary dataset: NYC Borough Boundaries shapefile for mapping purposes.
These datasets include a large number of observations, feature both temporal (date of complaint) and spatial (borough, latitude/longitude) fields, and provide several quantitative and categorical variables about the type of complaint, resolution time, and location.