**Rebecca Lersch**

**ISTA 421**

**December 4, 2025**

**Salena Ashton**

# Outline

1. **Introduction**
    a. **Discuss the Dataset of satellite information**
        i. **Features**
            1. **Official Name of Satellite**
            2. **Country/Organization of UN Registry**
            3. **Operator/Owner**
            4. **Country of Operator/Owner**
            5. **Users**
            6. **Purpose**
            7. **Detailed Purpose**
            8. **Class of Orbit**
            9. **Type of Orbit**
            10. **Longitude of Geosynchronous Orbit (Degrees)**
            11. **Perigee (Kilometers)**
            12. **Apogee (Kilometers)**
            13. **Eccentricity  Inclination (Degrees)**
            14. **Period (Minutes)**
            15. **Launch Mass (Kilograms)**
            16. **Dry Mass (Kilograms)**
            17. **Power (Watts)**
            18. **Date of Launch**
            19. **Expected Lifetime (Years)**
            20. **Contractor**
            21. **Country of Contractor**
            22. **Launch Site**
            23. **Launch Vehicle**
            24. **COSPAR Number**
            25. **NORAD Number**
    b. **Discuss the importance of the Dataset**
    c. **How different features are possible relevant**
        i. **Orbital regime, launch mass, lifetime, perigee, apogee, etc**
    d. **Identify the problem being solved, and the models being used**
        i. **Lifetime of satellite being predicted based on features**
2. **Why does this problem matter**
    a. **Reduce space trash if we can prolong the lifetime of a satellite**

        i. Satellites do not come back to earth, they become waste
            1. Can cause collisions
            2. Debris from collisions can fall back to earth or create more debris
    b. How is this information used
        i. How do engineers use these pieces of information
            1. Adjust aspects of satellites to prolong their life
    c. Background on satellite launches
        i. Types of satellites
        ii. Types of orbits
        iii. Types of missions
        iv. Why the mass might differ
        v. Why the rockets matter
        vi. Use my knowledge on this field here
            1. Use this for citations on this topic
                a. Cite my conference proceedings paper intro
    d. Background on different features that may be important
        i. GEO, LEO, NORAD, COSPAR, Perigee, Apogee
            1. What each of these do
            2. How they might be related
            3. Are they useful features
3. Predictions
    a. Which model will work better
    b. What features will be the most indicative
        i. Logistic
            1. Features
                a. Importance
        ii. Linear
            1. Features
                a. Importance
    c. Overall, what would this outcome mean
        i. Why do I think this is the case
4. Data preprocessing
    a. Exploration
    b. What features did I remove from the dataset
        i. NORAD
            1. ID number, incremental for all satellites launched since Sputnik, so not useful
        ii. COSPAR
            1. ID number, based off of date launched
        iii. Name
            1. Identification, can also change as different people acquire a satellite
        iv. Detailed Purpose
            1. A lot of blanks
            2. Would reduce the dataset too much to leave

3. A lot of variation
v. Country/Organization of UN Registry
1. Theres more detailed country information in other features

c. How did I clean
i. If orbital information is missing (period, apogee, perigee)
1. Use mean of satellites with same orbital class and fill
2. Explain why this is okay
a. Satellites in same orbital class are about close enough to each other it will be fine
ii. Date into datetime object
1. Useful for data exploration
iii. If no information on expected lifetime
1. Drop the row
iv. Split into two data frames
1. Linear regression data frame
2. Logistic regression data frame
v. Drop any other blank rows of data
vi. Split data frames into test and train

d. Visualizations of data

5. Linear Regression
a. Why did I choose
i. Have continuous variables
b. How did I set up
i. What features
1. Longitude of Geosynchronous Orbit (Degrees)
2. Perigee (Kilometers)
3. Apogee (Kilometers)
4. Eccentricity Inclination (Degrees)
5. Period (Minutes)
6. Launch Mass (Kilograms)
7. Dry Mass (Kilograms)
8. Power (Watts)
9. Expected Lifetime (Years)

c. What are the results
d. Visualization
e. Measure of fit

6. Linear regression interpretation
a. What do these results mean

7. Logistic regression
a. Why did I choose
i. Have discrete variables
b. How did I set up
i. What features
1. Operator/Owner
2. Country of Operator/Owner
3. Users

4. **Purpose**
5. **Class of Orbit**
6. **Type of Orbit**
7. **Date of Launch**
8. **Expected Lifetime (Years)**
9. **Contractor**
10. **Country of Contractor**
11. **Launch Site**
12. **Launch Vehicle**

   c. **What are the results**
   d. **Visualization**
   e. **Measure of fit**
8. **Logistic Regression Interpretation**
   a. **What do these results mean**
9. **Comparison of models**
   a. **Comparing the visual fit**
   b. **Comparing goodness of fit**
   c. **Difference between features used**
   d. **Which model is better**
      i. **Which feature is better**
10. **Conclusion**
   a. **Which model is better**
   b. **What does that mean for this (significance)**
   c. **Was my guess right?**
   d. **How does this extend outside of this problem**
11. **References**
   a. **Documentation**
   b. **Information on the field**

**Rubric (Where to find these points in my outline)**

**1. Problem Formulation and Significance (25 Points)**

- **Problem Statement (10 Points):** State a clear and concise description of the problem and your research question. Include the machine learning task(s) you've chosen and why it's important.
    - **Found in 1d**
- **Significance (10 Points):** Explain why this problem matters in your chosen field and how solving it contributes to real-world applications.
    - **Found in 1b**
- **Dataset Description (5 Points):** Detailed description of the dataset, including sources, qualitative and quantitative features, number of observations (at least 40), and why the dataset is suitable for your project.
    - **Found in 1a**

**2. Exploratory Data Analysis (15 Points)**

- **Data Exploration (5 Points):** Summary statistics and visualizations that explore both qualitative and quantitative features. Discuss any relationships or patterns found.
    - Found in 4a
- **Handling of Missing/Imbalanced Data (5 Points):** Explain how you handled missing, normalized, or imbalanced data.
    - **Found in 4b**
- **Data Visualization (5 Points):** Appropriate visualizations (e.g., histograms, boxplots, scatterplots) and meaningful interpretation of key data patterns and trends.
    - **Found in 4d**

**3. Model Selection, Application, and Evaluation (45 Points)**

- **Justify Model Choice (15 Points):** Explain the model(s) chosen with a clear rationale based on dataset characteristics and your research question/ problem statement. Model choice needs to reflect what was learned in class (linear models, tree-based methods, etc.).
    - **Found in 1d, 5a, 7a**
- **Method Applications (15 Points):** Correctly implement your model(s) and include appropriate hyperparameter tuning. If you choose a non-parametric model, you must include the process of choosing your metrics (e.g., how you chose k for KNN or clustering).
    - **Found in 5b, 7b**
- **Model Evaluation (15 Points):** Evaluate your model performance using relevant metrics (e.g., accuracy, precision, recall, AUC, RMSE). Compare results across models or hyperparameters and discuss the trade-offs involved.
    - **Found in 5e, 7e, 9**

**4. Results, Conclusions, and Real-World Implications (30 Points)**

- **Results Presentation (10 Points):** Present your results in a clear manner using tables, plots, and key statistics. Results should be tied back to the problem statement and dataset.
    - **Found in 10a**
- **Conclusions (10 Points):** Interpret the results, discussing key takeaways and any limitations. Discussion should be thoughtful and relate to the problem objectives.
    - **Found in 10**

- **Workforce/Graduate School Preparation (10 Points):** Explain how your project developed or sharpened your skills that are relevant for future goals, whether for entering the workforce or applying to graduate school. Specific skills (e.g., data analysis, programming, optimization, focusing on the question at hand, communication, etc... ) should be discussed.
    - **Found in 10d**