Day 1: Probability Distribution of Survival Rates

- **Problem Description:**
- Calculate the probability distribution of survival rates across passenger classes using the Titanic dataset. Visualize the survival probabilities for each class as a bar chart.

- **Mathematical Formulation:**
- P(S|C) = Number of survivors in class C / Total passengers in class C

- **Steps to Solve:**
- 1. Group the dataset by passenger class (C).
- 2. Compute the number of survivors and total passengers in each class.
- 3. Calculate P(S|C) for each class.
- 4. Visualize the results as a bar chart.

- **Dataset**
- url = 'https://raw.githubusercontent.com/datasciencedojo/datasets/master/titanic.csv'

Day 2: Correlation Analysis Using Heatmaps

- **Problem Description:**
- Compute and visualize the correlation matrix between delay durations and numerical features like flight distance using the Airline Delay dataset.

- **Mathematical Formulation:**
- Correlation coefficient: $r_{xy} = (\Sigma(x_i - \bar{x})(y_i - \bar{y})) / \sqrt{(\Sigma(x_i - \bar{x})^2 \Sigma(y_i - \bar{y})^2)}$

- **Steps to Solve:**
- 1. Select numerical features from the dataset.
- 2. Calculate the pairwise correlation coefficients (r_xy) for all feature pairs.
- 3. Visualize the resulting correlation matrix as a heatmap.

- **Dataset**
- https://www.kaggle.com/datasets/patrickzel/flight-delay-and-cancellation-dataset-2019-2023

Day 3: Moving Average of Flight Delays

- **Problem Description:**
- Calculate the 7-day moving average of total flight delays per day and visualize the trend.

- **Mathematical Formulation:**
- 7-day moving average: $MA_t = (\Sigma x_i) / 7$, for $t \geq 7$, where x_i represents total delays on day i.

- **Steps to Solve:**
- 1. Aggregate the total delays per day.
- 2. Compute the 7-day moving average.
- 3. Plot the original daily delays and the moving average as a line chart.

- **Dataset**
- https://www.kaggle.com/datasets/patrickzel/flight-delay-and-cancellation-dataset-2019-2023

Day 4: Distribution of Delays by Distance

- **Problem Description:**
- Compute the distribution of delays by flight distance intervals (e.g., short-haul, medium-haul, long-haul) and create an interactive visualization.

- **Mathematical Formulation:**
- Average delay for interval I_k: $\bar{D}_k = (\Sigma D(d)) / \text{Number of flights in } I_k$, where $d \in I_k$.

- **Steps to Solve:**
- 1. Define flight distance intervals.
- 2. Group flights into intervals and calculate the average delay ($\bar{D}_k$) for each.
- 3. Create an interactive visualization using Plotly with dropdown filters.

- **Dataset**
- url = 'https://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data'

Day 5: Linear Regression for Passenger Satisfaction

- **Problem Description:**
- Build a linear regression model to predict overall satisfaction scores based on features like flight distance, delays, and service quality. Visualize the regression line and residuals.

- **Mathematical Formulation:**
- Linear regression: $\hat{y} = \beta_o + \beta_1 x_1 + \beta_2 x_2 + ... + \beta_n \quad x_n$, where $\hat{y}$ is the predicted satisfaction score.

- **Steps to Solve:**
- 1. Select features and target variable (satisfaction scores).
- 2. Fit a linear regression model using the dataset.
- 3. Visualize the regression line for a single feature.
- 4. Create a residual plot to analyze the model's performance.

- **Dataset**
- https://www.kaggle.com/code/ahmedabbas757/airline-passenger-satisfaction-eda/input