1. Figure out a nice dataset and submit a project proposal by the end of 9.30 pm.  
2. Data cleaning - Panadas  
2. Data analysis - Pandas  
4. Visualization - Mathplolib or pandas  (6- 8) figure .  
5. Summary of your findings.  
6. Push everything to your GitHub repo- read me and .

**Project Proposal (10 Marks)**

* Identify an interesting and relevant dataset.
* Provide a clear justification and rationale for dataset selection.
* Outline well-defined project objectives and methodology.
* Submit the proposal on time (by the specified deadline).

**Data Cleaning with Pandas (20 Marks)**

* Identify and handle missing values.
* Remove duplicates and outliers.
* Ensure correct data types and consistency.
* Provide clear, well-documented code explaining the cleaning process.

**Data Analysis with Pandas (20 Marks)**

* Conduct thorough exploratory data analysis (EDA).
* Demonstrate effective use of Pandas for data manipulation.
* Interpret and explain the analytical results in relation to project objectives.

**Visualization (Matplotlib/Pandas) (20 Marks)**

* Create between 6 to 8 meaningful figures.
* Ensure all visualizations are clearly labeled (titles, axes, legends).
* Choose appropriate graph types to represent the data.
* Use design elements (color, layout) effectively to enhance clarity.

**Summary of Findings (10 Marks)**

* Provide a concise and clear summary of key insights.
* Organize the summary logically.
* Clearly connect the findings back to the original project objectives.

**GitHub Repository & Documentation (10 Marks)**

* Maintain a well-organized repository with all relevant code and resources.
* Include a comprehensive README file that outlines project details, setup instructions, and usage.

**Teamwork (10 Marks)**

* Demonstrate equal contribution from all team members.
* Show effective collaboration and communication.

Your project on road accidents in the UK sounds well-structured! Here are a few suggestions to enhance your analysis:

1. **Data Preprocessing:**
   * Check for missing or inconsistent values in the dataset.
   * Convert date columns into a datetime format for better analysis.
   * Categorize weather, lighting, and road surface conditions into meaningful groups if needed.
2. **Exploratory Data Analysis (EDA):**
   * Use histograms and box plots to visualize accident trends.
   * Create heatmaps to show correlations between different factors.
3. **Statistical Analysis:**
   * Use Chi-square tests to check for significant relationships.
   * Conduct regression analysis to identify key contributing factors.
4. **Visualization Techniques:**
   * Time series plots for accident trends.
   * Bar charts comparing accidents in urban vs. rural areas.
   * Pie charts or stacked bar charts for casualties by month.

Let me know if you need help with Python code, dataset handling, or specific visualizations!

1. Figure out a nice dataset and submit a project proposal by the end of 9.30 pm.  
2. Data cleaning - Panadas  
2. Data analysis - Pandas  
4. Visualization - Mathplolib or pandas  (6- 8) figure .  
5. Summary of your findings.  
6. Push everything to your GitHub repo- read me and .

**Project Proposal (10 Marks)**

* Identify an interesting and relevant dataset.
* Provide a clear justification and rationale for dataset selection.
* Outline well-defined project objectives and methodology.
* Submit the proposal on time (by the specified deadline).

**Data Cleaning with Pandas (20 Marks)**

* Identify and handle missing values.
* Remove duplicates and outliers.
* Ensure correct data types and consistency.
* Provide clear, well-documented code explaining the cleaning process.

**Data Analysis with Pandas (20 Marks)**

* Conduct thorough exploratory data analysis (EDA).
* Demonstrate effective use of Pandas for data manipulation.
* Interpret and explain the analytical results in relation to project objectives.

**Visualization (Matplotlib/Pandas) (20 Marks)**

* Create between 6 to 8 meaningful figures.
* Ensure all visualizations are clearly labeled (titles, axes, legends).
* Choose appropriate graph types to represent the data.
* Use design elements (color, layout) effectively to enhance clarity.

**Summary of Findings (10 Marks)**

* Provide a concise and clear summary of key insights.
* Organize the summary logically.
* Clearly connect the findings back to the original project objectives.

**GitHub Repository & Documentation (10 Marks)**

* Maintain a well-organized repository with all relevant code and resources.
* Include a comprehensive README file that outlines project details, setup instructions, and usage.

**Teamwork (10 Marks)**

* Demonstrate equal contribution from all team members.
* Show effective collaboration and communication.

Your project on road accidents in the UK sounds well-structured! Here are a few suggestions to enhance your analysis:

1. **Data Preprocessing:**
   * Check for missing or inconsistent values in the dataset.
   * Convert date columns into a datetime format for better analysis.
   * Categorize weather, lighting, and road surface conditions into meaningful groups if needed.
2. **Exploratory Data Analysis (EDA):**
   * Use histograms and box plots to visualize accident trends.
   * Create heatmaps to show correlations between different factors.
3. **Statistical Analysis:**
   * Use Chi-square tests to check for significant relationships.
   * Conduct regression analysis to identify key contributing factors.
4. **Visualization Techniques:**
   * Time series plots for accident trends.
   * Bar charts comparing accidents in urban vs. rural areas.
   * Pie charts or stacked bar charts for casualties by month.

Let me know if you need help with Python code, dataset handling, or specific visualizations!