Performing EDA In Education Sector for Engg

Sprint 1-2: Project Planning and Data Collection

Objectives:

- Define the project scope and objectives.
- Identify and collect relevant data sources.

Tasks:

1. **Project Scope Definition:**
    - Define the specific questions you want to answer (e.g., trends in engineering admissions, pass rates, demographic analysis).
    - Outline the deliverables and success criteria.
2. **Data Source Identification:**
    - Identify sources of data such as government databases (e.g., All India Survey on Higher Education), university websites, and relevant educational bodies.
3. **Data Collection:**
    - Collect data on engineering colleges, courses offered, student demographics, admission statistics, pass rates, faculty details, etc.
    - Ensure data coverage for multiple years for trend analysis.

Sprint 3-4: Data Preprocessing and Cleaning

Objectives:

- Clean and preprocess the collected data to ensure consistency and accuracy.

Tasks:

1. **Data Cleaning:**
    - Handle missing values and inconsistencies.
    - Remove duplicates and irrelevant data.
    - Standardize data formats (e.g., dates, numeric values).
2. **Data Integration:**
    - Merge datasets from different sources.
    - Ensure consistent keys (e.g., college codes, course codes).
3. **Data Validation:**
    - Validate data against known benchmarks or additional sources.

- Perform initial exploratory checks for anomalies or outliers.

Sprint 5-6: Exploratory Data Analysis (Initial Insights)

Objectives:

- Conduct initial exploratory data analysis to uncover basic insights.

Tasks:

1. **Descriptive Statistics:**
    - Compute summary statistics (mean, median, mode, standard deviation).
    - Analyze distribution of key variables (e.g., number of colleges, student demographics).
2. **Initial Visualizations:**
    - Create basic visualizations (histograms, box plots, bar charts).
    - Identify patterns or trends (e.g., enrollment over the years, gender ratio).
3. **Preliminary Insights:**
    - Document initial findings and potential areas of interest for deeper analysis.

Sprint 7-8: Advanced EDA and Feature Extraction

Objectives:

- Perform deeper exploratory analysis and extract relevant features for further analysis.

Tasks:

1. **Correlation Analysis:**
    - Analyze relationships between variables (e.g., correlation between admission rates and pass rates).
2. **Advanced Visualizations:**
    - Create more complex visualizations (heatmaps, scatter plots, trend lines).
    - Use interactive visualizations for better insights (e.g., dashboards).
3. **Feature Extraction:**
    - Identify and extract key features for further analysis (e.g., college reputation, location impact, faculty qualifications).

Sprint 9-10: Hypothesis Testing and Insights Refinement

Objectives:

- Test hypotheses and refine insights based on EDA results.

Tasks:

1. **Hypothesis Formulation:**
    - Develop hypotheses based on initial findings (e.g., Does college location affect admission rates?).

2. **Hypothesis Testing:**
    - Conduct statistical tests (e.g., t-tests, chi-square tests) to validate hypotheses.
    - Analyze test results and refine insights.
3. **Insights Documentation:**
    - Document validated insights and any interesting anomalies or trends.
    - Prepare a detailed report on findings.

Sprint 11-12: Data Storyline Development

Objectives:

- Develop a coherent data storyline to present findings effectively.

Tasks:

1. **Storyline Creation:**
    - Structure the insights into a compelling narrative.
    - Highlight key findings, trends, and actionable insights.
2. **Visualization Enhancement:**
    - Enhance visualizations for better storytelling (e.g., annotations, highlights).
    - Create a storyboard with key visuals and narratives.
3. **Presentation Preparation:**
    - Prepare a presentation deck.
    - Include both visual and textual content to convey insights clearly.

Sprint 13-14: Reporting and Documentation

Objectives:

- Compile the final report and document the entire analysis process.

Tasks:

1. **Final Report Compilation:**
    - Compile all findings, visualizations, and narratives into a comprehensive report.
    - Ensure the report is well-structured and easy to understand.
2. **Documentation:**
    - Document the data sources, cleaning process, and analysis methods.
    - Include code snippets and methodologies for reproducibility.
3. **Review and Feedback:**
    - Share the report with stakeholders for feedback.

- Incorporate feedback and finalize the report.

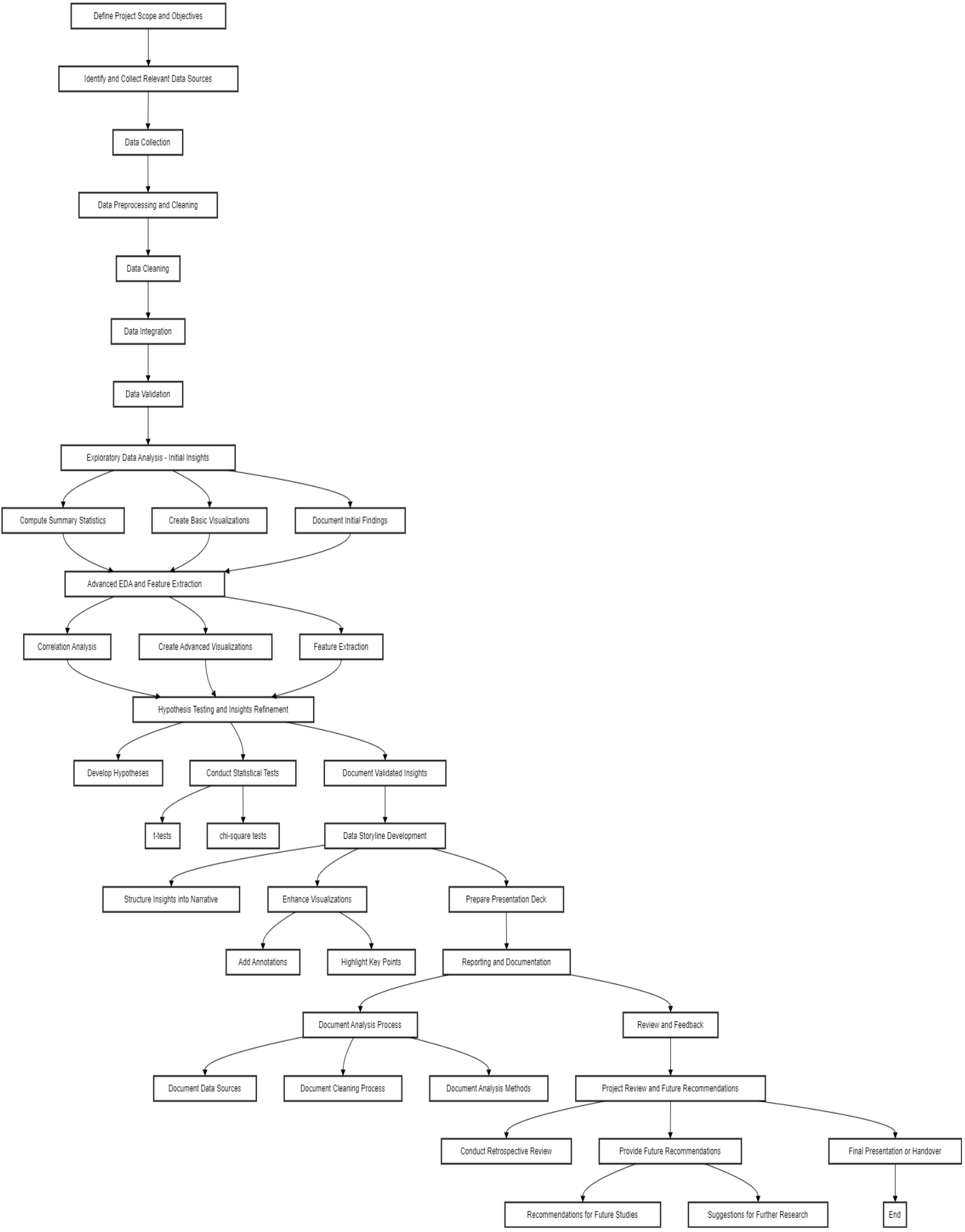
Sprint 15-16: Project Review and Future Recommendations

Objectives:

- Review the project and provide recommendations for future work.

Tasks:

1. **Project Review:**
    - Conduct a retrospective review of the project.
    - Identify what went well and areas for improvement.
2. **Future Recommendations:**
    - Provide recommendations for future studies or data collection.
    - Suggest potential areas for further research based on findings.
3. **Closure:**
    - Finalize all project documentation.
    - Conduct a final presentation or handover session with stakeholders.


Define Project Scope and Objectives
Identify and Collect Relevant Data Sources
Data Collection
Data Preprocessing and Cleaning
Data Cleaning
Data Integration
Data Validation
Exploratory Data Analysis - Initial Insights
Compute Summary Statistics
Create Basic Visualizations
Document Initial Findings
Advanced EDA and Feature Extraction
Correlation Analysis
Create Advanced Visualizations
Feature Extraction
Hypothesis Testing and Insights Refinement
Develop Hypotheses
Conduct Statistical Tests
Document Validated Insights
t-tests
chi-square tests
Data Storyline Development
Structure Insights into Narrative
Enhance Visualizations
Prepare Presentation Deck
Add Annotations
Highlight Key Points
Reporting and Documentation
Document Analysis Process
Review and Feedback
Document Data Sources
Document Cleaning Process
Document Analysis Methods
Project Review and Future Recommendations
Conduct Retrospective Review
Provide Future Recommendations
Final Presentation or Handover
Recommendations for Future Studies
Suggestions for Further Research
End

Explanation of the Activity Diagram:

1. **Define Project Scope and Objectives**:
    - Start by defining the project scope and objectives.
2. **Identify and Collect Relevant Data Sources**:
    - Identify and collect the relevant data sources needed for the analysis.
3. **Data Collection**:
    - Collect data from identified sources.
4. **Data Preprocessing and Cleaning**:
    - Preprocess and clean the collected data.
5. **Data Cleaning**:
    - Handle missing values, remove duplicates, and standardize data formats.
6. **Data Integration**:
    - Merge datasets and ensure consistent keys.
7. **Data Validation**:
    - Validate the data against benchmarks and perform initial exploratory checks.
8. **Exploratory Data Analysis - Initial Insights**:
    - Perform initial exploratory data analysis to gain insights.
9. **Compute Summary Statistics**:
    - Calculate summary statistics of key variables.
10. **Create Basic Visualizations**:
    - Create basic visualizations like histograms, box plots, and bar charts.
11. **Document Initial Findings**:
    - Document the initial findings from the analysis.
12. **Advanced EDA and Feature Extraction**:
    - Perform advanced exploratory data analysis and extract key features.
13. **Correlation Analysis**:
    - Analyze relationships between variables.
14. **Create Advanced Visualizations**:
    - Create advanced visualizations such as heatmaps, scatter plots, and trend lines.
15. **Feature Extraction**:
    - Identify key features from the data.

16. **Hypothesis Testing and Insights Refinement**:
    - Develop hypotheses and conduct statistical tests to refine insights.
17. **Develop Hypotheses**:
    - Formulate hypotheses based on initial findings.
18. **Conduct Statistical Tests**:
    - Perform statistical tests like t-tests and chi-square tests.
19. **Document Validated Insights**:
    - Document the validated insights from hypothesis testing.
20. **Data Storyline Development**:
    - Develop a data storyline to present insights effectively.
21. **Structure Insights into Narrative**:
    - Structure the insights into a cohesive narrative.
22. **Enhance Visualizations**:
    - Enhance visualizations with annotations and highlights.
23. **Prepare Presentation Deck**:
    - Prepare the presentation deck for reporting.
24. **Reporting and Documentation**:
    - Compile the final report and document the analysis process.
25. **Document Analysis Process**:
    - Document the data sources, cleaning process, and analysis methods.
26. **Review and Feedback**:
    - Review the work and gather feedback.
27. **Project Review and Future Recommendations**:
    - Conduct a retrospective review and provide future recommendations.
28. **Conduct Retrospective Review**:
    - Reflect on the project process and outcomes.
29. **Provide Future Recommendations**:
    - Provide recommendations for future studies and research.
30. **Final Presentation or Handover**:
    - Conduct the final presentation or handover the project deliverables.
31. **End**:

- End of the process.