# A Comparative Analysis of Traditional Data Analysis Methods Versus LLM-Assisted Approaches on the Titanic Dataset

# Abstract

This paper presents a detailed comparison of traditional data analysis methods and LLM-assisted approaches using the Titanic dataset. We demonstrate the efficiency and accuracy of both approaches in data cleaning, exploratory data analysis, and feature engineering. The results suggest that LLMs can significantly expedite routine tasks while providing valuable insights comparable to those obtained through traditional methods.

# 1. Introduction

Data analysis has rapidly evolved with advancements in technology, and Large Language Models (LLMs) like GPT-4 have begun to play a pivotal role in automating data analysis tasks. This paper evaluates the strengths and limitations of traditional data analysis techniques versus LLM-assisted methods using the Titanic Passenger Dataset, emphasizing their respective roles in a modern data workflow.

# 2. Literature Review

## 2.1 Traditional Data Analysis Methods

Traditional data analysis often involves using programming libraries such as Pandas, NumPy, and visualization tools like Matplotlib. It requires significant expertise and time, particularly for tasks like data cleaning, handling missing values, and exploratory data analysis (EDA). Despite being rigorous and accurate, traditional methods can be time-consuming, especially when dealing with large or unstructured datasets.

## 2.2 LLM-Assisted Data Analysis

LLMs such as GPT-4 have shown immense potential in automating data analysis tasks, from data cleaning to generating insightful visualizations. Researchers have found that LLMs can identify patterns and trends effectively, thus enhancing the efficiency of data analysis workflows. The ability to understand natural language prompts makes LLMs a powerful tool for analysts, bridging the gap between complex data science tasks and human understanding.

# 3. Methodology

## 3.1 Data Selection

We used the Titanic Passenger Dataset, containing 891 rows and 12 columns with a mix of numerical and categorical data, to compare the efficiency and effectiveness of traditional and LLM-assisted analysis approaches.

## 3.2 Traditional Data Analysis Approach

Data Cleaning:  
  
Checked for missing values in the dataset using Pandas.  
Handled missing values by filling numerical values (e.g., 'Age') with the median and categorical values (e.g., 'Embarked') with the mode.  
Detected missing values:  
'Age': 177 missing values  
'Cabin': 687 missing values  
'Embarked': 2 missing values

Code Example:

# Checking for missing values  
print(titanic\_data.isnull().sum())  
  
# Handling missing values  
titanic\_data['Age'].fillna(titanic\_data['Age'].median(), inplace=True)  
titanic\_data['Embarked'].fillna(titanic\_data['Embarked'].mode()[0], inplace=True)

## 3.3 LLM-Assisted Data Analysis Approach

The LLM suggested handling missing values based on relationships within the dataset, such as using group means for 'Age' based on 'Pclass' and 'Sex'.

Code Example:

# Fill missing 'Age' values using group mean  
titanic\_data['Age'] = titanic\_data['Age'].fillna(titanic\_data.groupby(['Pclass', 'Sex'])['Age'].transform('mean'))

# 4. Experiments and Results

Both approaches were independently conducted by the same analyst to ensure consistency. The time taken for each task, accuracy of insights, and usability were recorded.

The following table summarizes the time taken by both approaches:

|  |  |  |
| --- | --- | --- |
| Task | Traditional Approach (Time) | LLM-Assisted Approach (Time) |
| Data Cleaning | 45 minutes | 15 minutes |
| Exploratory Data Analysis | 1 hour | 30 minutes |
| Statistical Analysis | 1.5 hours | 45 minutes |

# 5. Discussion

The LLM-assisted approach significantly enhanced the efficiency of the data analysis process. It not only expedited routine tasks but also uncovered additional insights that might have been overlooked in the traditional approach.

# 6. Conclusion

This comparative study demonstrates that LLMs offer considerable benefits in enhancing efficiency, accuracy, and insights within data analysis workflows. However, human analysts remain essential for validating results and ensuring data quality. Integrating LLMs with traditional data analysis processes can lead to more effective and insightful analyses.

# 7. Recommendations for Future Work

Exploring the integration of LLMs with more advanced machine learning models for predictive analytics.  
Investigating the use of LLMs for real-time data analysis, such as financial market trends or social media sentiment.