Formula 1 Race Position Performance- Stock Price Analysis

Milestone 1: Data Collection, Preprocessing, and Exploratory Data Analysis

Project Objective:

To develop an interactive dashboard that analyzes Formula 1 race performance metrics and their potential correlation with team market performance. The project aims to uncover patterns in race performance metrics and lay the groundwork for future analysis of how these metrics might influence team company stock price.

Technical Stack:

- **Programming Language**: Python 3.x
- **Key Libraries**:
  - Data Manipulation: pandas, numpy
  - Visualization: matplotlib, seaborn
  - Statistical Analysis: scipy
  - Machine Learning (future): scikit-learn
  - Dashboard Development (future): Streamlit
- **Development Environment**:
  - IDE: Jupyter Notebook
  - Version Control: Git/GitHub
  - Documentation: Markdown

Dataset Description

The analysis uses the Formula 1 World Championship dataset, which includes:

- Races data (1950-2024)
- Driver information
- Lap times
- Pit stops
- Constructor details
- Each Team's Stock Price in the race period

Data set link can be found here.

Data Preprocessing Steps

1. **Missing Data Handling**
   - Numerical values: Median imputation

- o Categorical values: Mode imputation
2. **Outlier Detection and Treatment**
    - o Z-score method for lap times
    - o IQR method for pit stop durations
3. **Feature Engineering**
    - o Created derived metrics for performance analysis
    - o Standardized time-based features

## Key Insights from EDA

1. **Lap Time Analysis**
    - o Notable improvement in lap times from 2000-2004, reaching fastest period (~87-88 seconds)
    - o Significant slowdown from 2008-2010, peaking around 108 seconds in 2010
    - o Gradual improvement post-2010 with some fluctuations
    - o Recent years (2020-2024) show relatively stable lap times around 95-98 seconds
    - o The variations likely correspond to major regulation changes and technological developments in F1


Distribution of Lap Times

Average Lap Times by Year
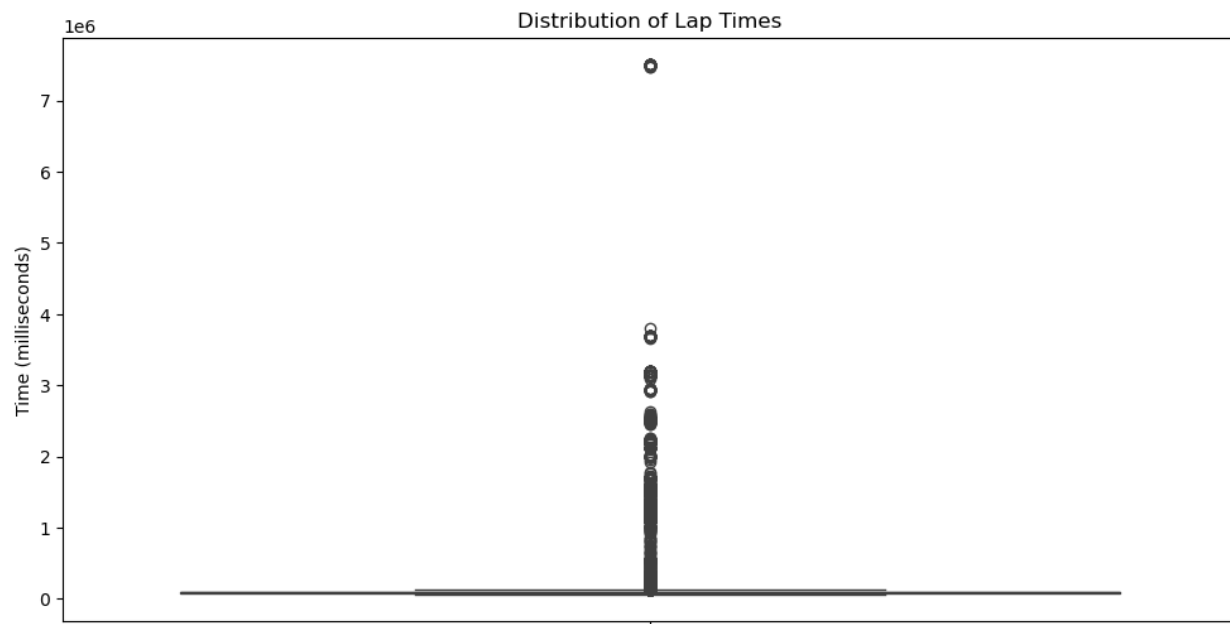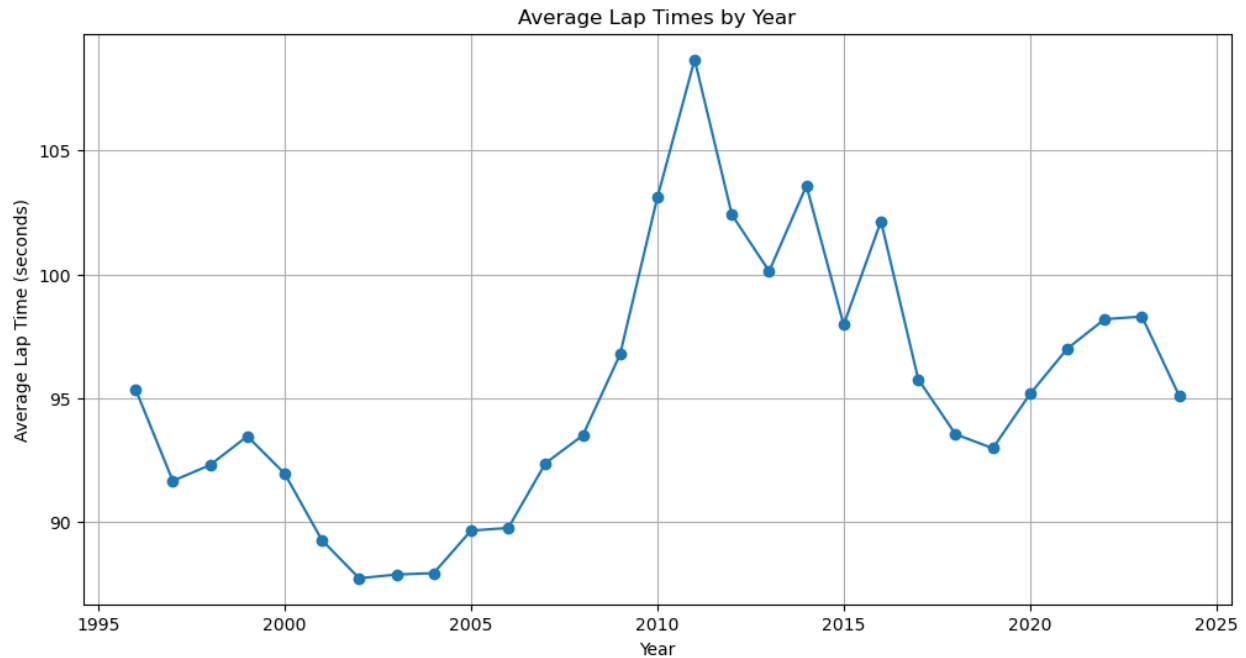
2. **Position Analysis**
   o Clear negative correlation between average position and position variation
   o Drivers with better average positions (lower numbers, 4-8) show higher position variation (3-7 std), indicating more dynamic racing and overtaking at the front of the grid
   o Back-markers (positions 16-20) show much lower position variation (1-3 std), suggesting they tend to remain in similar positions
   o Middle-field positions (10-14) show moderate variation, indicating competitive mid-pack racing
   o The scatter pattern suggests three distinct racing tiers: front-runners, mid-field, and back-markers.


Position Consistency Analysis
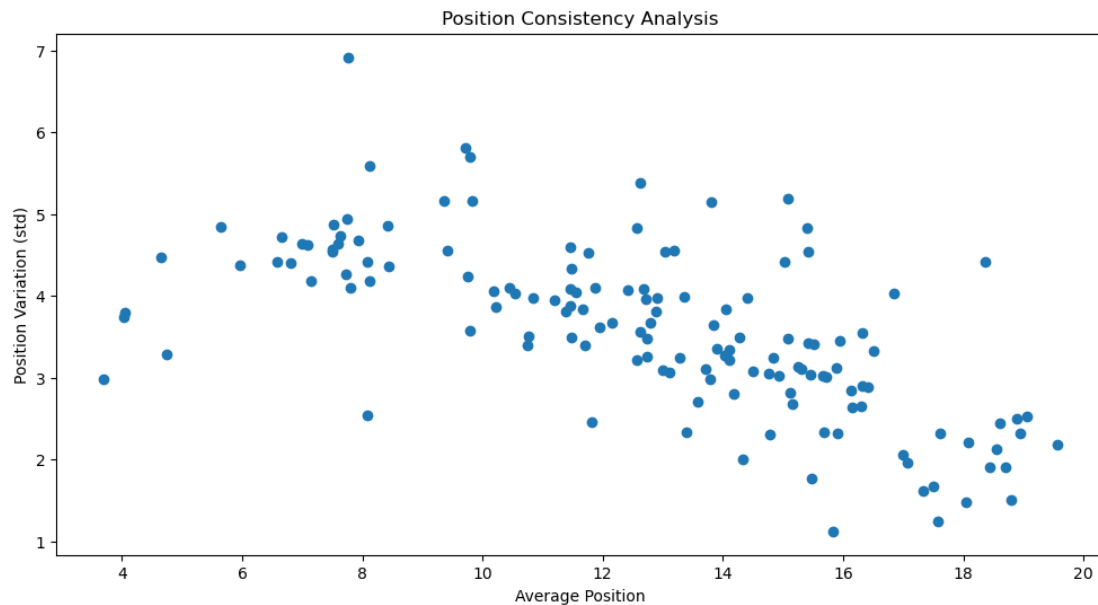
3. **Driver Performance Patterns**
   o British drivers dominate Formula 1 with approximately 160 drivers, followed closely by American drivers
   o There's a significant drop after the top 2 nationalities, with Italian drivers being the third most common
   o European nations (British, French, German, Belgian, Swiss) make up the majority of the top 10
   o South African representation shows F1's global reach, though at a much lower count
   o The distribution suggests a historical European-centric nature of the sport, particularly British dominance.

Top 10 Driver Nationalities
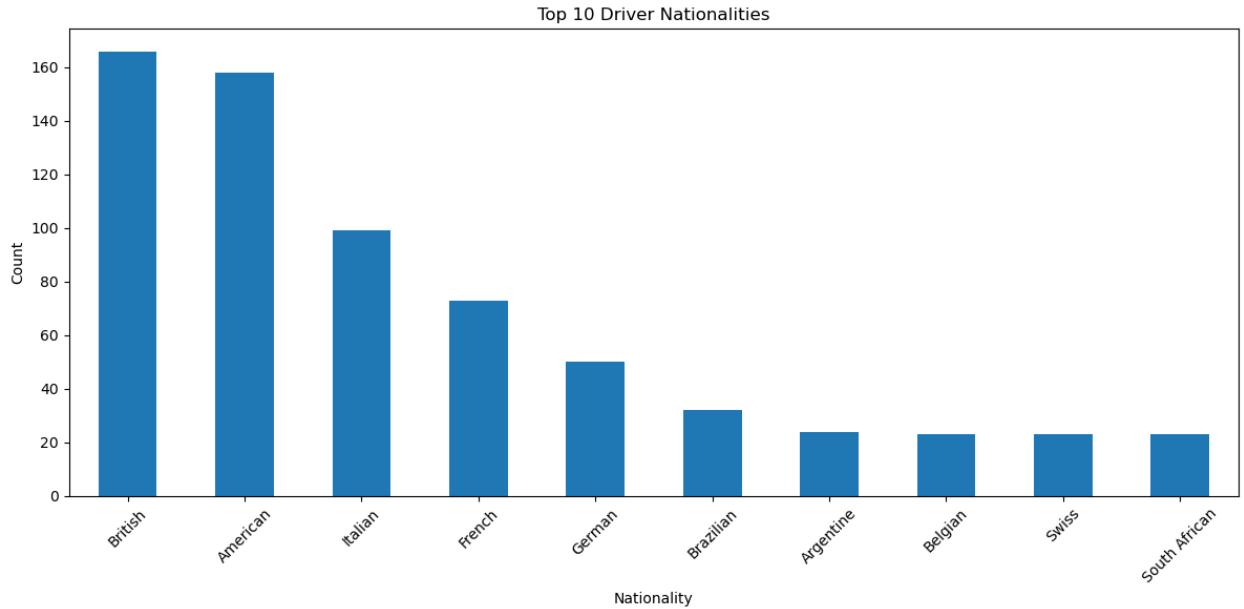
4. **Constructor Performance**
   o Ferrari leads significantly with over 10,000 total points
   o Mercedes and Red Bull form a clear second tier with ~7,500 points each
   o McLaren stands alone in a third tier with ~7,000 points
   o Significant drop-off after the top 4 teams
   o Historical teams like Williams and Renault show decent point accumulation
   o Newer or smaller teams (Force India, Team Lotus, Benetton, Tyrrell) have significantly fewer points
   o The distribution suggests a clear hierarchy in F1 team success and resources

Top 10 Constructors by Total Points
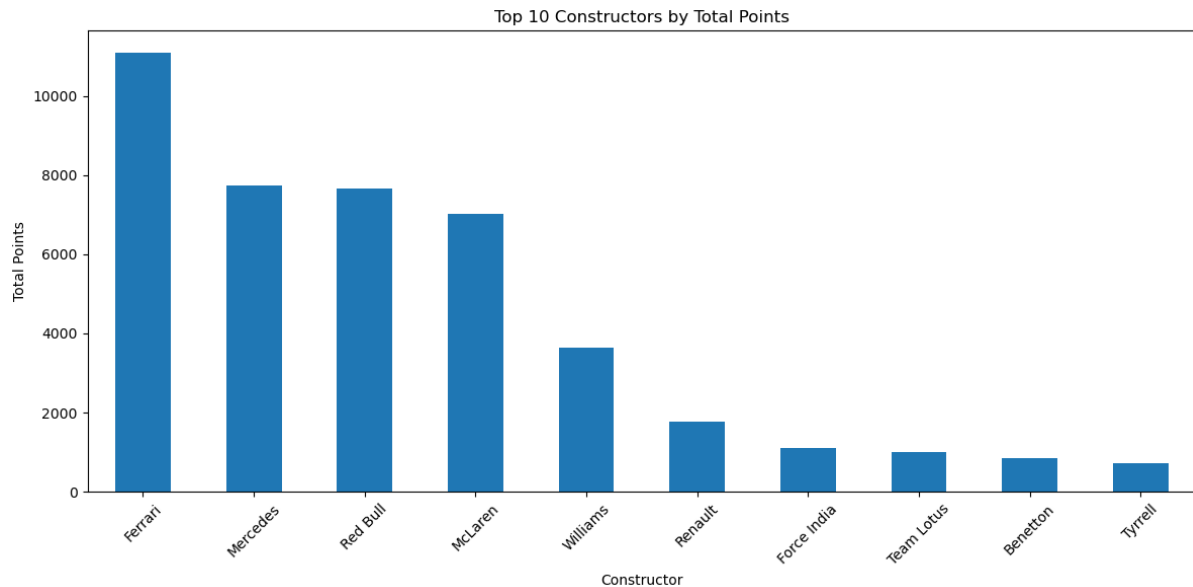
Project Timeline

1. **Milestone 1** (Data Collection & EDA) - February 5- February 21, 2025
2. **Milestone 2** (Advanced feature engineering, Model development for performance prediction & Initial dashboard structure) - February 21- March 15, 2025
3. **Milestone 3** (Dashboard development, Performance optimization & Final documentation and presentation) - March 15-April 15, 2025

Next Steps

- Integration of financial market data
- Development of advanced performance metrics
- Creation of predictive models
- Initial dashboard prototype development