CSE 564 Visualization - Prelim Report

Driven to Break: Visualizing Failures and Reliability in Formula 1

Team 20

Harsh Vivek Londhekar ID: #116647641
Aitik Dandapat ID: #116626453

📊 Dataset Overview

The project continues to leverage the F1DB (Formula 1 Database), encompassing comprehensive Formula 1 historical racing data from the community-maintained open-source dataset. The dataset remains an integral resource for analyzing mechanical reliability, driver performance, and strategic factors affecting race outcomes.

⚙️ Data Preparation - Completed

1. Data Merging

We successfully merged multiple relevant tables (race metadata, driver details, constructor details, finishing statuses, and failure causes) to form a unified master dataset enabling comprehensive and efficient analysis.

2. Data Cleaning

The merged data underwent rigorous cleaning, including removal of duplicates, normalization of categorical fields, and consolidation of inconsistent labels for failure causes, ensuring accurate and effective analyses.

3. Data Sampling

We maintained a balanced sampling strategy using stratified sampling based on decades and teams, ensuring fair representation and unbiased insights.

4. Feature Selection

Selected key features such as year, circuit ID, driver ID, constructor ID, grid position, laps completed, finish position, and retirement reasons remain integral. Derived features such as the DNF_flag and FailureType have proven essential for highlighting critical reliability and safety insights.

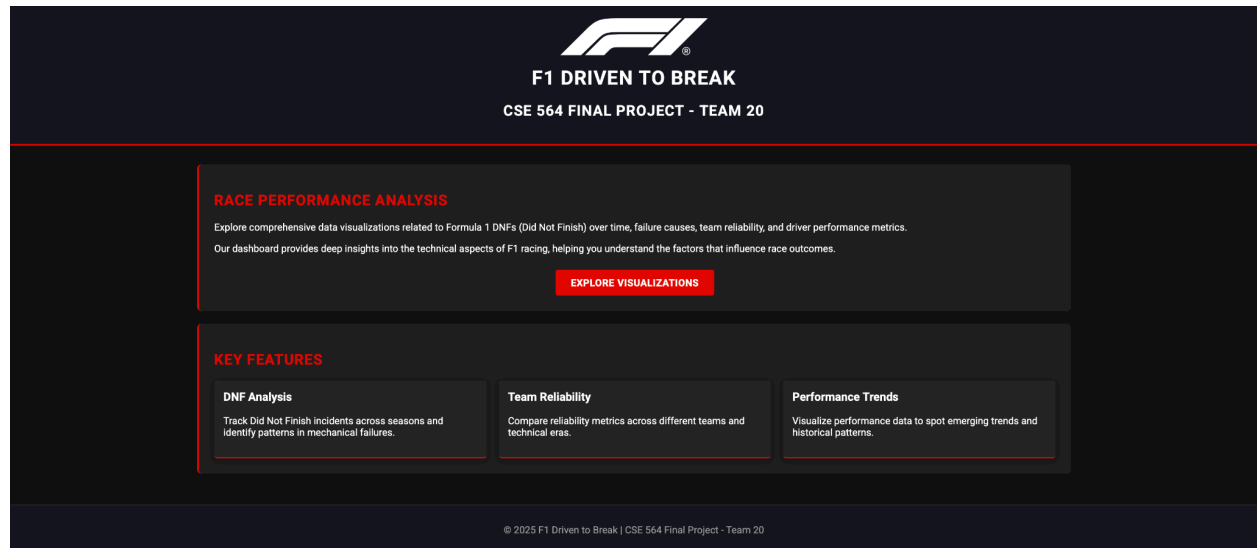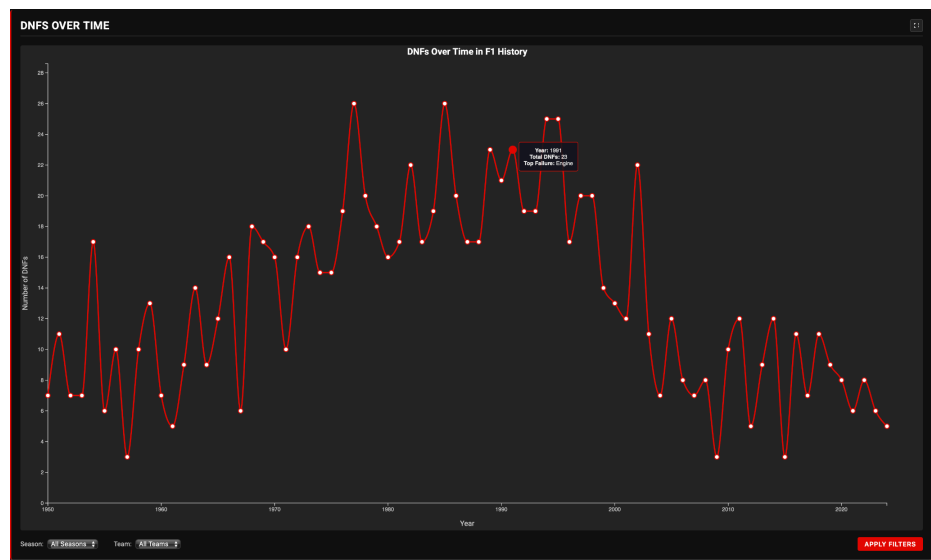# 📈 Progress on Visualizations - In Progress

Fig 1: Home Page

Fig 2: Main DashBoard

1. DNFs Over Time (Line Plot)

The line plot visualizing DNFs across historical F1 seasons effectively highlights periods of peak mechanical unreliability. Interactive elements such as tooltips, zoom/pan, and brushing have been implemented, enabling linked interactions with other visualizations for detailed analysis.

2. Failure Cause Breakdown (Treemap)
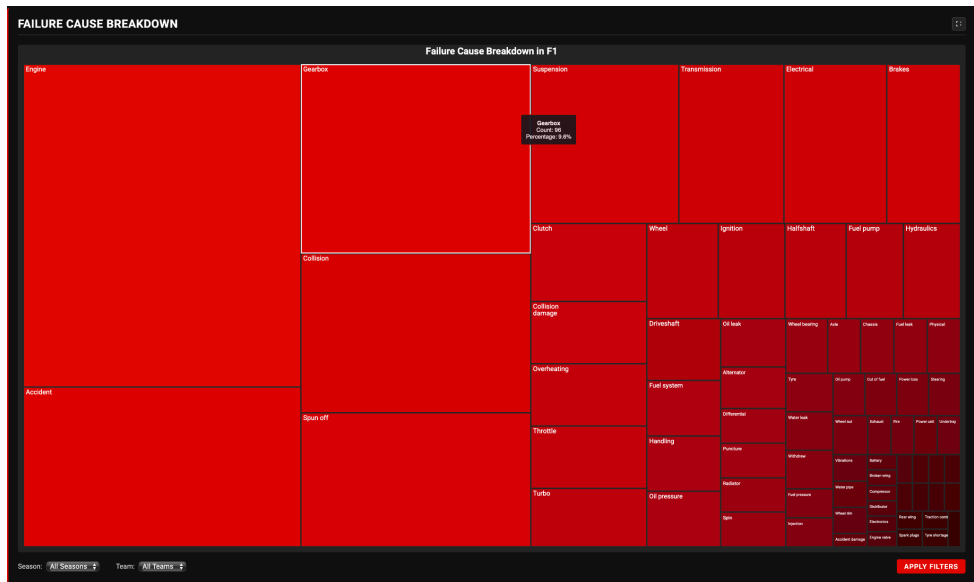
The treemap has effectively illustrated the proportional distribution of various DNF causes such as mechanical failures, collisions, and technical issues. Interactive filtering, selection, and linking to other visualizations enhance user engagement and analytical depth.

3. Team Reliability Ranking (Stacked Bar Chart)
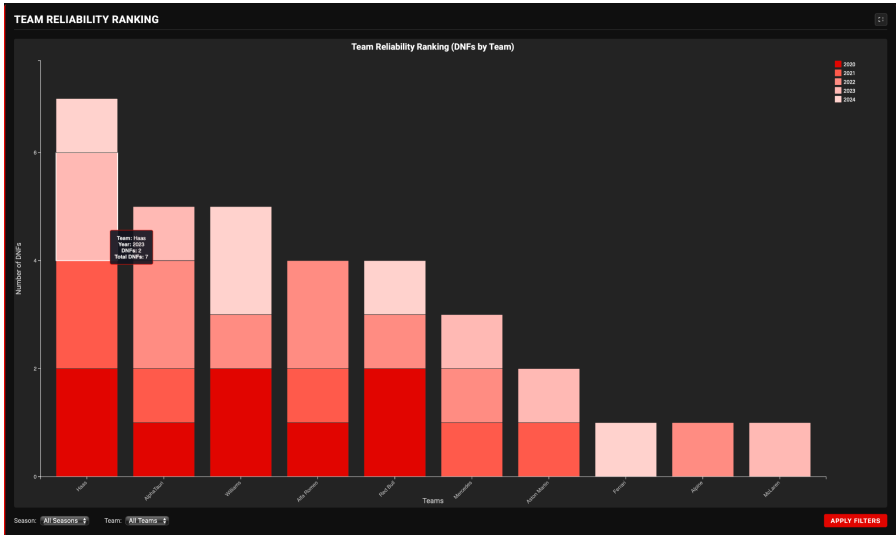
This stacked bar chart enables effective visualization of team reliability over recent seasons, clearly demonstrating team-specific trends in reliability and failure occurrences. Interactive brushing and linking features allow users to dynamically explore and cross-reference data.

### 4. Tyre vs Engine Failures (Grouped Bar Chart)
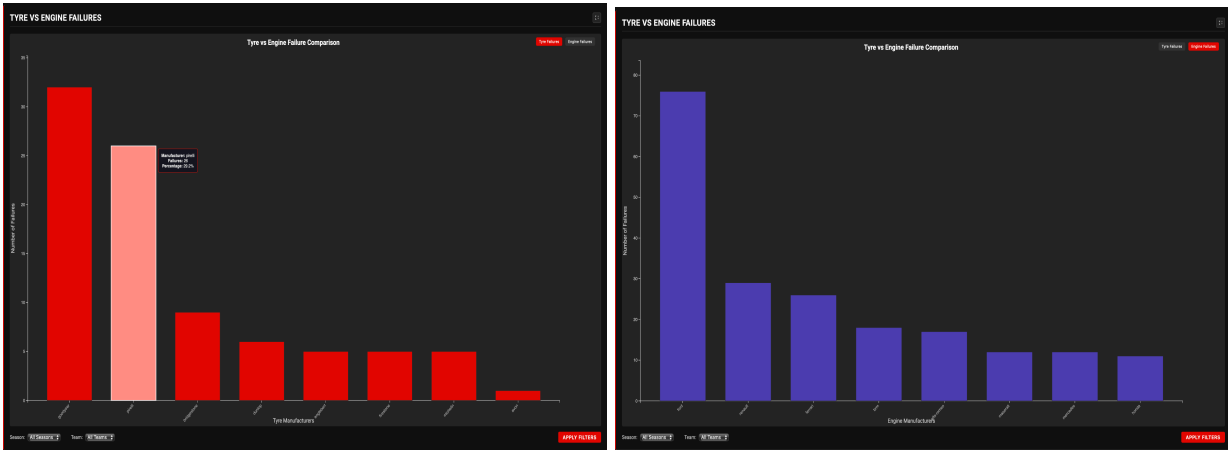
Interactive grouped bar charts effectively compare reliability issues between tyre manufacturers and engine manufacturers. Users can toggle views between tyre-related and engine-related failures, with linked interactions enabling comparative insights across visualizations.

5. Driver Experience vs DNFs (Scatter Plot)

The scatter plot successfully correlates driver experience (number of races completed) with their DNF ratios, effectively identifying potential experience-related trends. Interactive scatter plot is under development and will incorporate brushing and linking features for comprehensive comparative analyses.

6. Circuit Risk Index (Geo Map Area Chart)
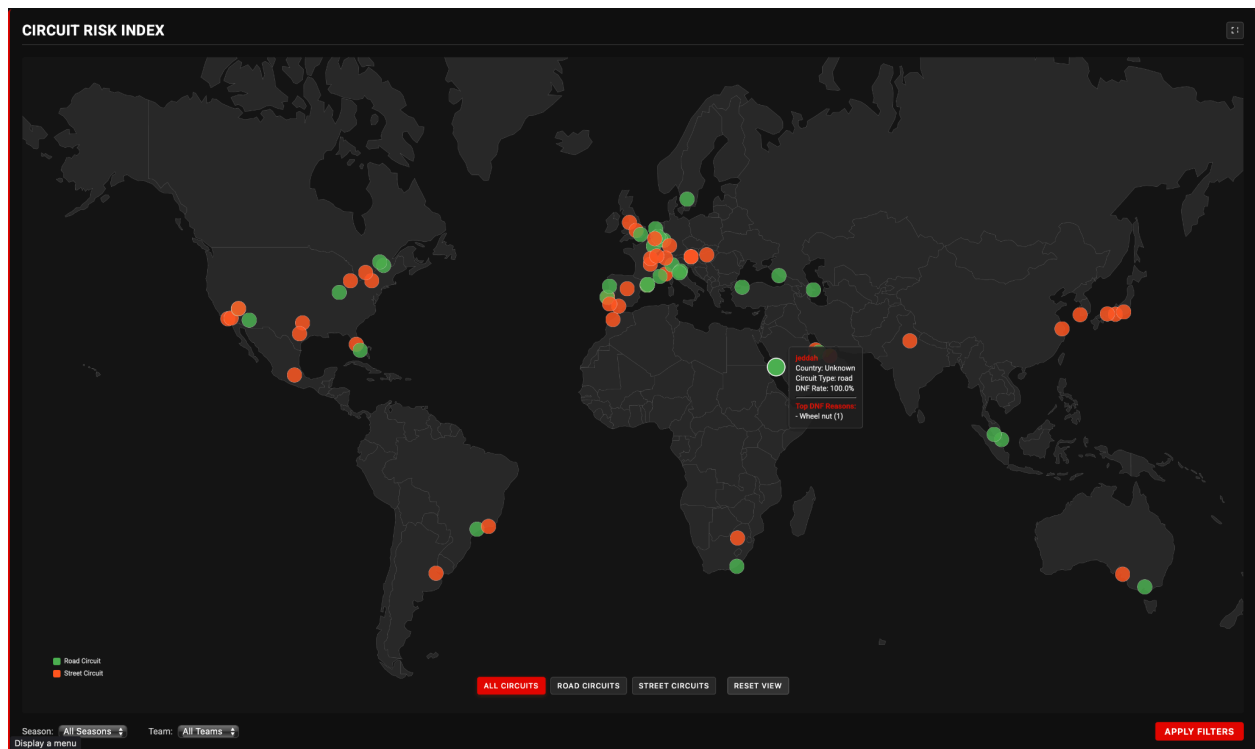
The geo map area chart effectively highlights the geographic distribution of high-risk circuits, clearly differentiating between road and street circuits. Interactive tooltips provide details such as DNF ratios and top reasons for failures, with brushing and linking functionality allowing users to correlate geographic insights with other data.

7. Factors Influencing DNFs: Multivariable Analysis (Parallel Coordinate Plot)

It displays attributes like grid position, laps, pit stops, fastest lap speed, and a DNF flag to explore multi-variable patterns. This plot is under development.

🎯 Next Steps

- Finalize interactive components and visual optimizations.
- Complete and incorporate the Driver Experience vs DNFs (Scatter Plot).
- Complete and incorporate the Factors Influencing DNFs (PCP Plot)
- Enhance overall dashboard usability and visual aesthetics.