F1

MACHINE LEARNING

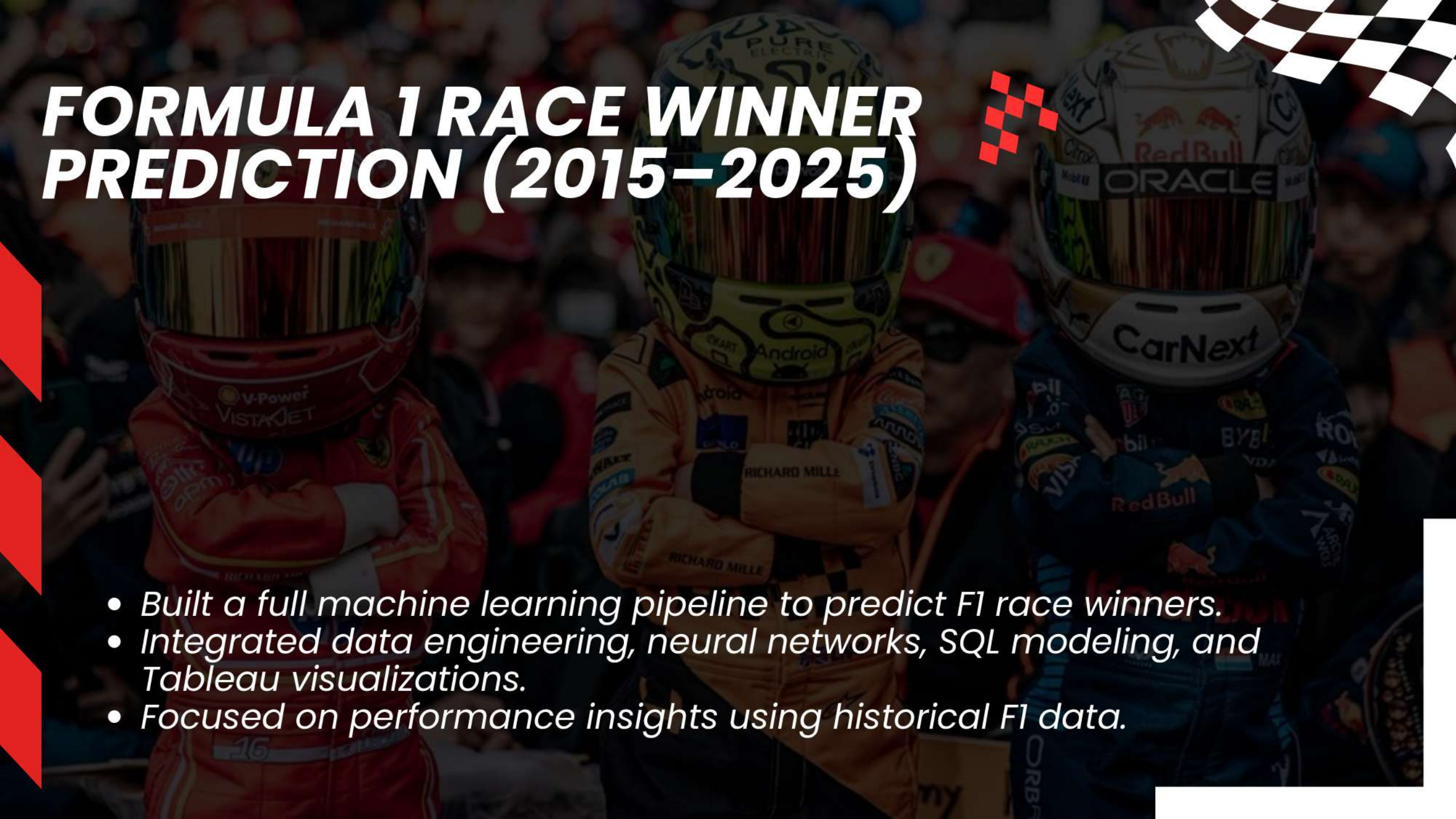
Data Analysis Bootcamp: Project 4

TEAM

- *Dariana Ibarra*
- *Jose Manuel Romero Oceguera*
- *Miguel Angel Olmos Valderrama*
- *Valeria Itzel Jimenez Paz*

INDEX

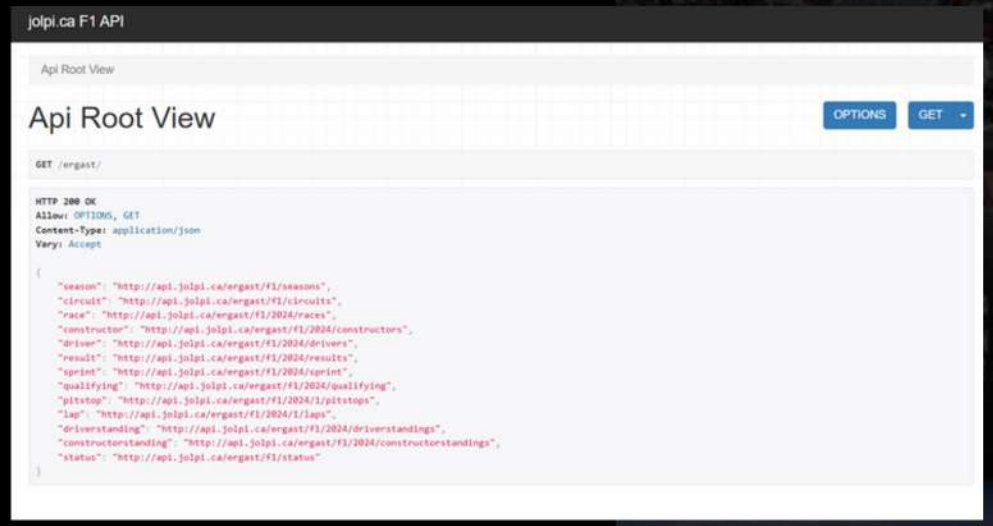
FORMULA 1 RACE WINNER PREDICTION (2015–2025)



- *Built a full machine learning pipeline to predict F1 race winners.*
- *Integrated data engineering, neural networks, SQL modeling, and Tableau visualizations.*
- *Focused on performance insights using historical F1 data.*

DATA COLLECTION FROM JOLPI.CA API

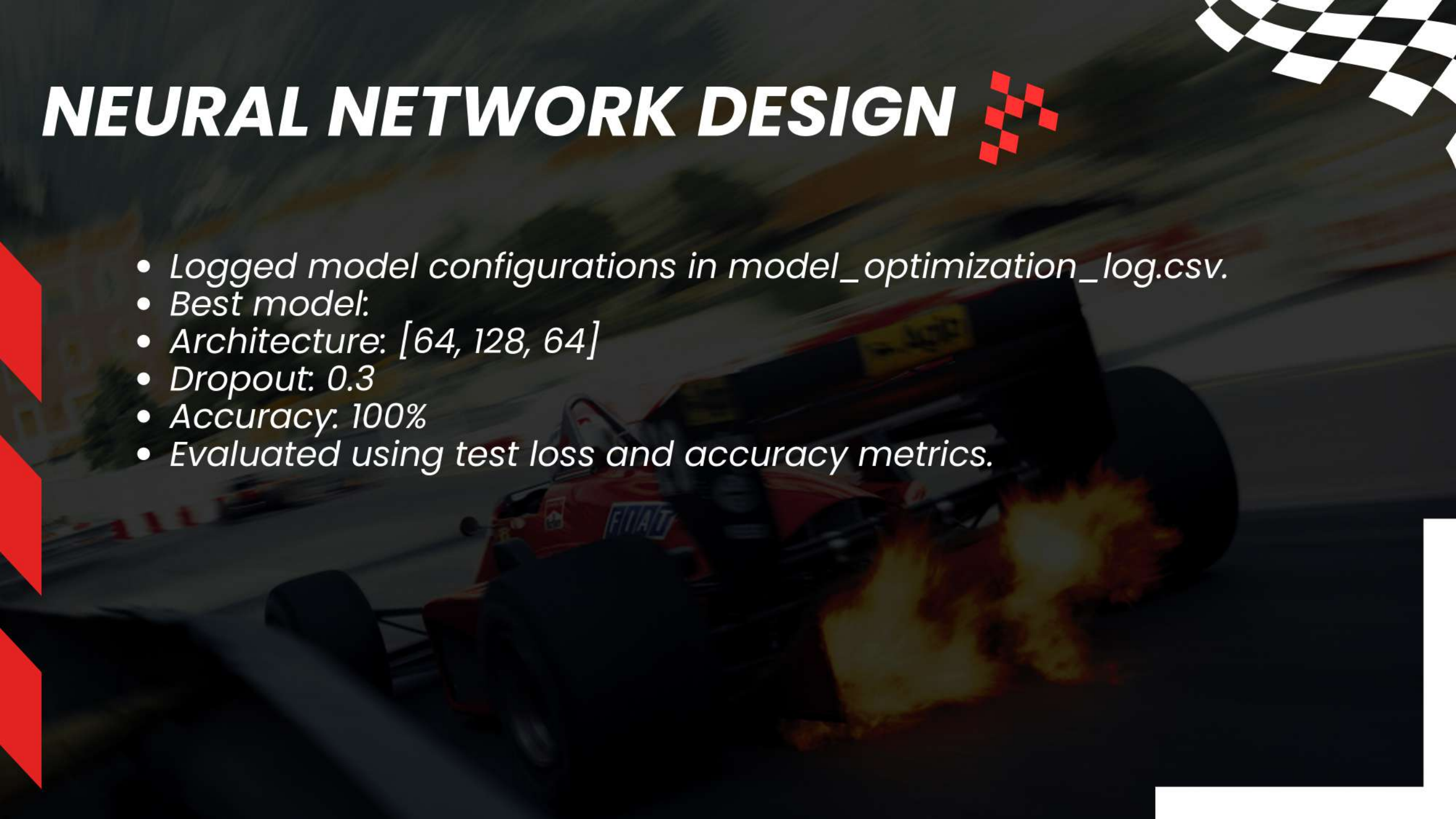
- *Dynamic race data pulled from:*
- *https://api.jolpi.ca/ergast/f1/{season}/results.json*
- *Included fields: season, round, driver, constructor, position.*
- *Cleaned and structured for modeling.*



ETL PROCESS & FEATURE PREP

- *Removed nulls and irrelevant columns.*
- *Encoded winner names using LabelEncoder.*
- *Scaled season and round using StandardScaler.*

NEURAL NETWORK DESIGN



- *Logged model configurations in model_optimization_log.csv.*
- *Best model:*
- *Architecture: [64, 128, 64]*
- *Dropout: 0.3*
- *Accuracy: 100%*
- *Evaluated using test loss and accuracy metrics.*

PREDICTING THE NEXT WINNER

- *Input: 2025 – Round 1*
- *Output: Piastri*
- *Validated that prediction pipeline works with new input.*

```
🏁 Final Best Model Summary:
Model ID: 1
Architecture: [64, 128, 64]
Dropout Rate: 0.3
Learning Rate: 0.001
Batch Size: 42
Test Accuracy: 1.0000
Test Loss: 0.6673
1/1 ━━━━━━━━━━━━━━━━━━━━ 0s 344ms/step

🏆 Predicted Winner for 2025 Round 1: Piastri
```

POSTGRESQL DATA MODEL

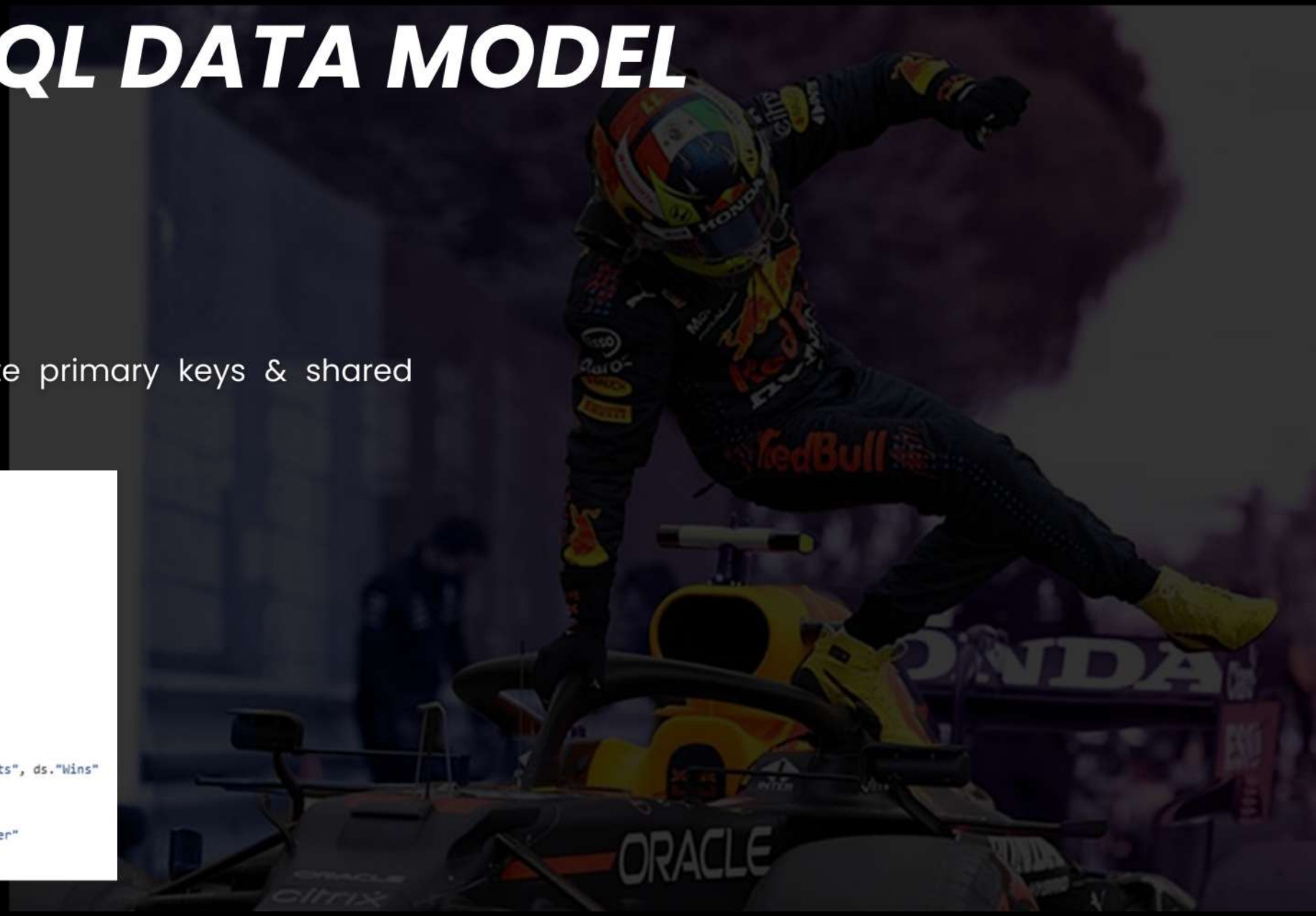
Tables created:

- constructor_standings
- driver_standings
- qualifying_results
- race_results
- Designed with composite primary keys & shared columns for joins.

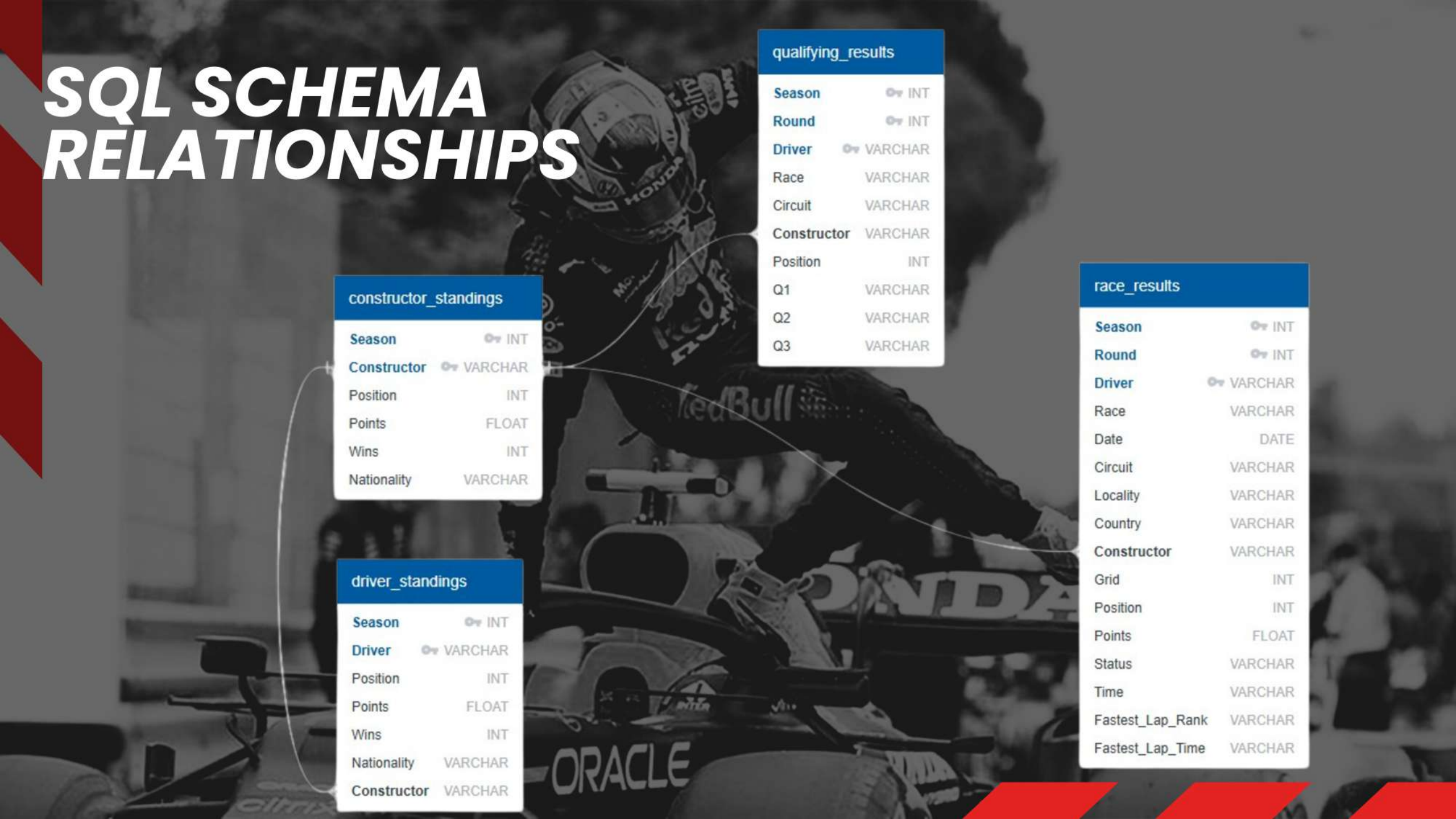
```sql
-- Top constructors by average points
SELECT "Constructor", AVG("Points") AS avg_points
FROM constructor_standings
GROUP BY "Constructor"
ORDER BY avg_points DESC;

-- Drivers with most race wins
SELECT "Driver", COUNT(*) AS wins
FROM race_results
WHERE "Position" = 1
GROUP BY "Driver"
ORDER BY wins DESC;

SELECT r."Season", r."Driver", r."Constructor", r."Points", ds."Wins"
FROM race_results r
JOIN driver_standings ds
  ON r."Season" = ds."Season" AND r."Driver" = ds."Driver"
LIMIT 10;
```



SQL SCHEMA RELATIONSHIPS



qualifying_results

Season	INT
Round	INT
Driver	VARCHAR
Race	VARCHAR
Circuit	VARCHAR
Constructor	VARCHAR
Position	INT
Q1	VARCHAR
Q2	VARCHAR
Q3	VARCHAR

constructor_standings

Season	INT
Constructor	VARCHAR
Position	INT
Points	FLOAT
Wins	INT
Nationality	VARCHAR

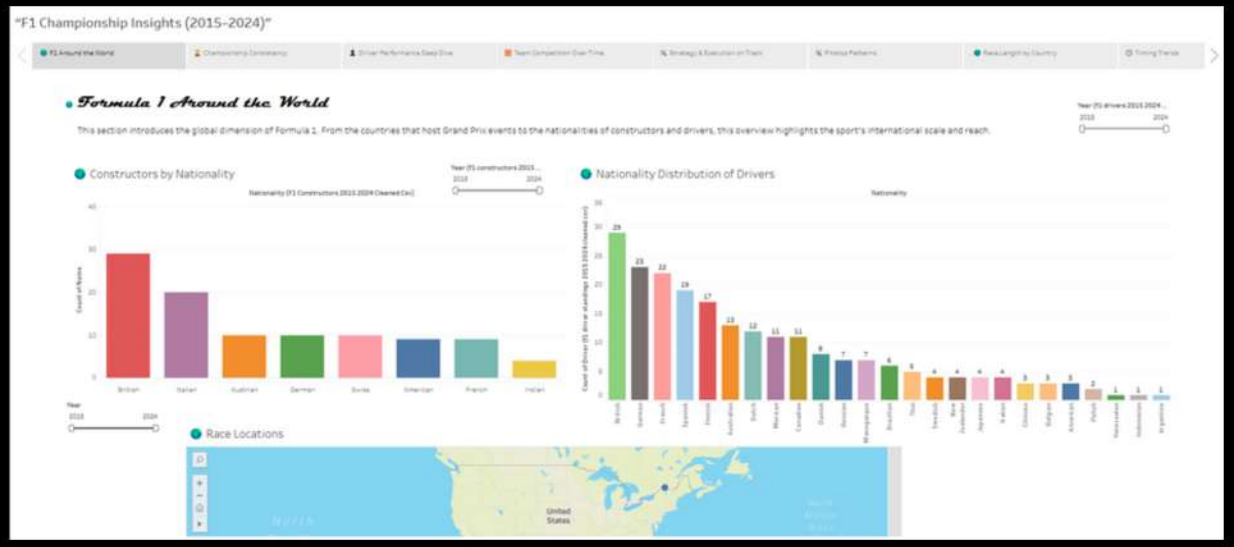
race_results

Season	INT
Round	INT
Driver	VARCHAR
Race	VARCHAR
Date	DATE
Circuit	VARCHAR
Locality	VARCHAR
Country	VARCHAR
Constructor	VARCHAR
Grid	INT
Position	INT
Points	FLOAT
Status	VARCHAR
Time	VARCHAR
Fastest_Lap_Rank	VARCHAR
Fastest_Lap_Time	VARCHAR

driver_standings

Season	INT
Driver	VARCHAR
Position	INT
Points	FLOAT
Wins	INT
Nationality	VARCHAR
Constructor	VARCHAR

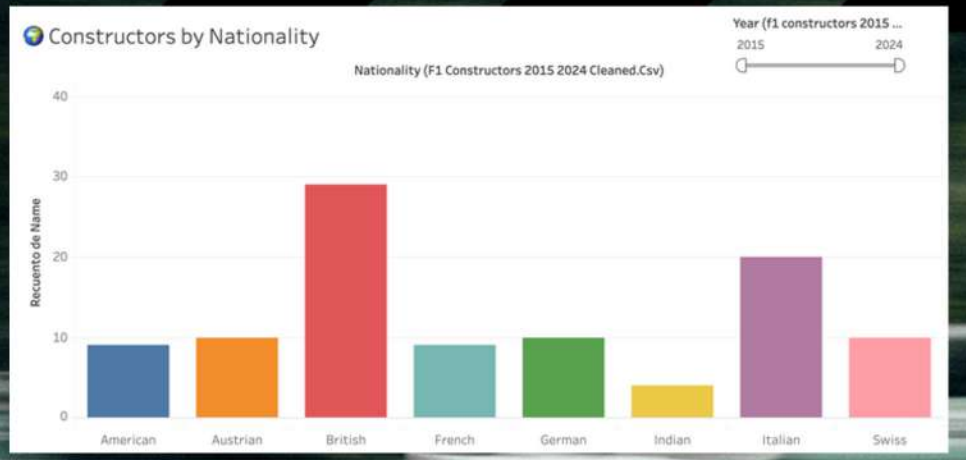
F1 INSIGHTS – TABLEAU DASHBOARD

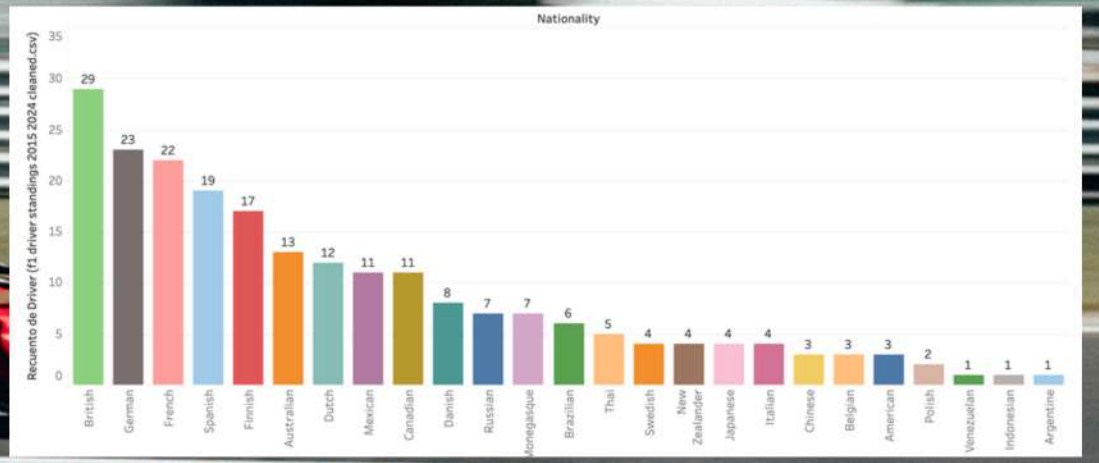


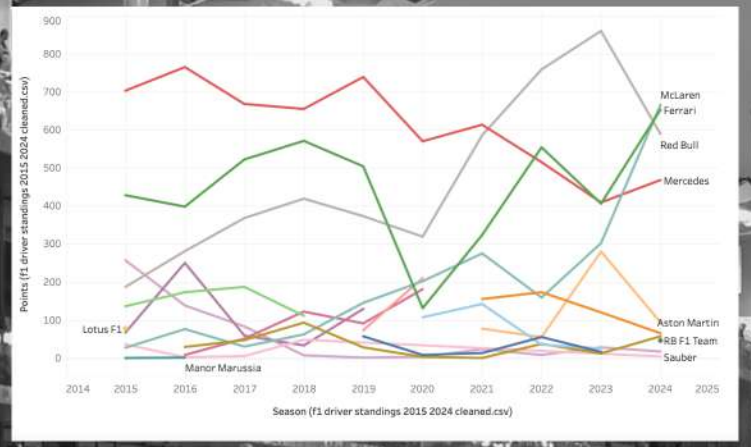
- *Race duration (longest/shortest by country)*
- *Pit stop durations by driver*
- *Driver and constructor performance*
- *Qualifying vs race outcome comparison*

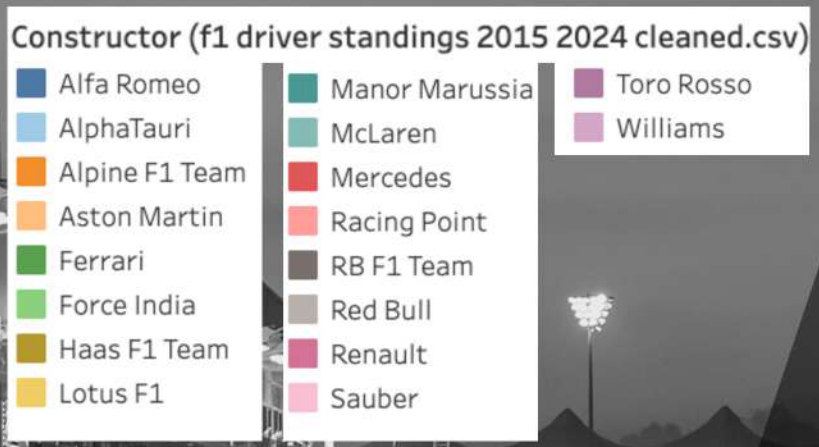
F1 AROUND THE WORLD

- Constructors by Nationality
- Nationality Distribution of Drivers
- Race Locations





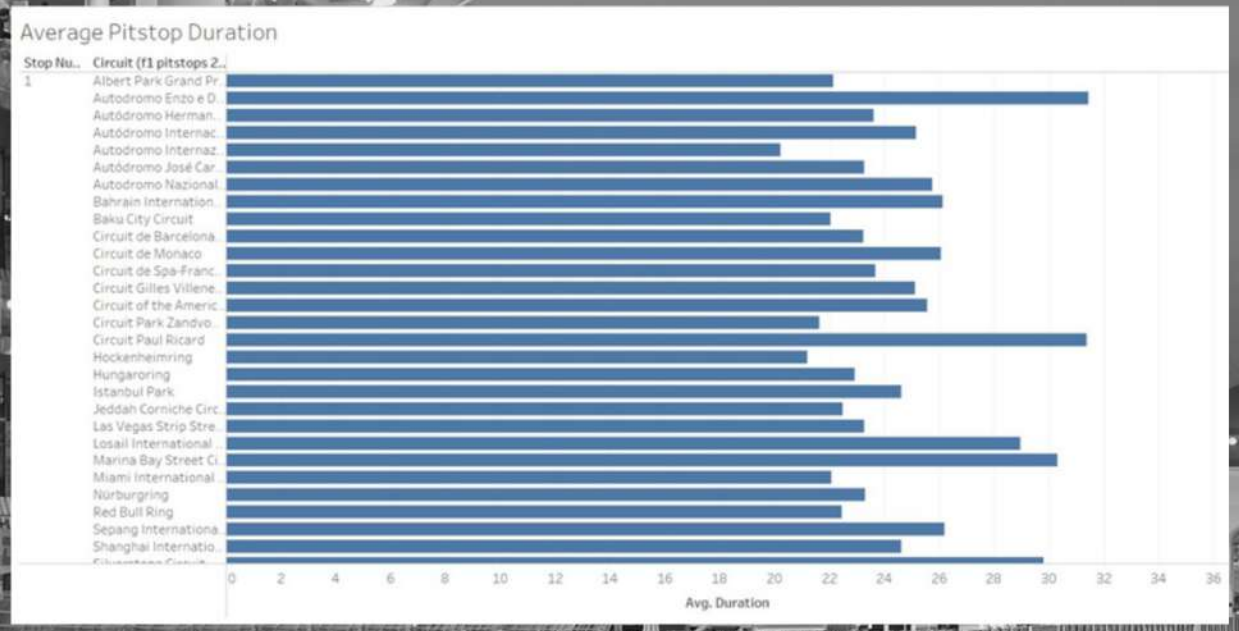


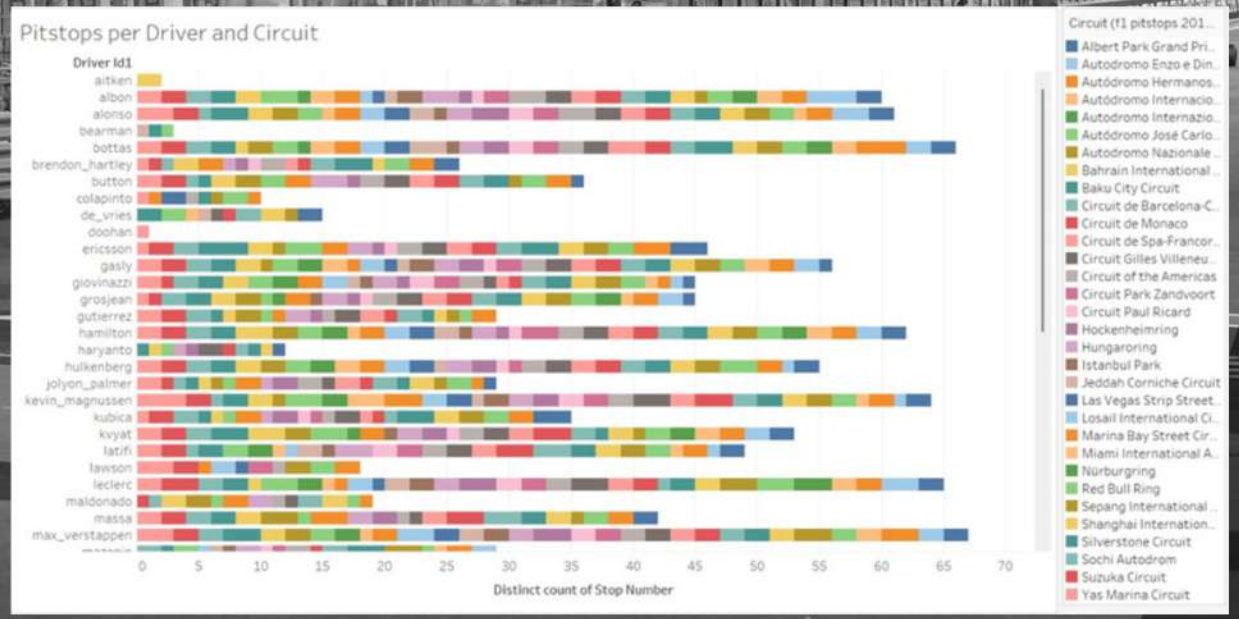


- Position Rank by Driver
- Championship Points by Constructor

F1'S TOP PERFORMERS

F1 PITSTOP ANALYSIS

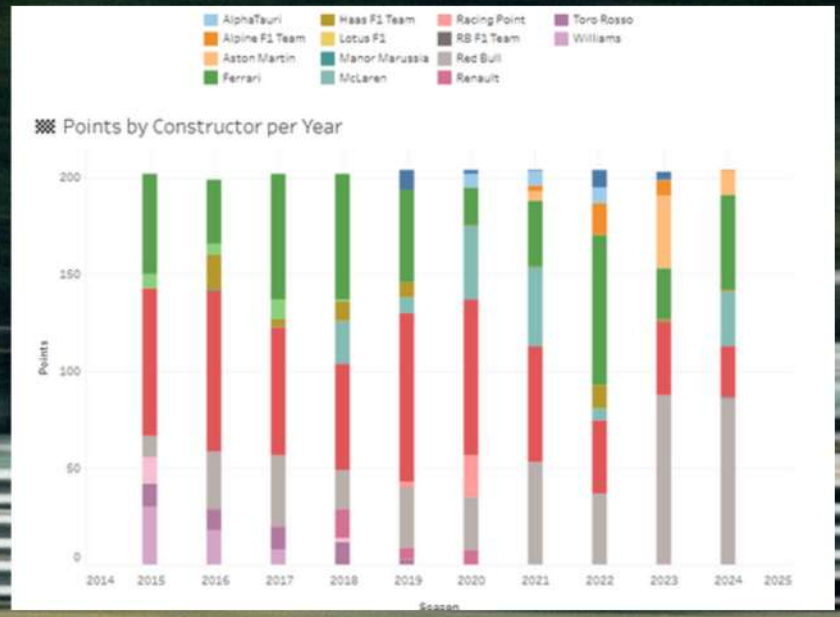


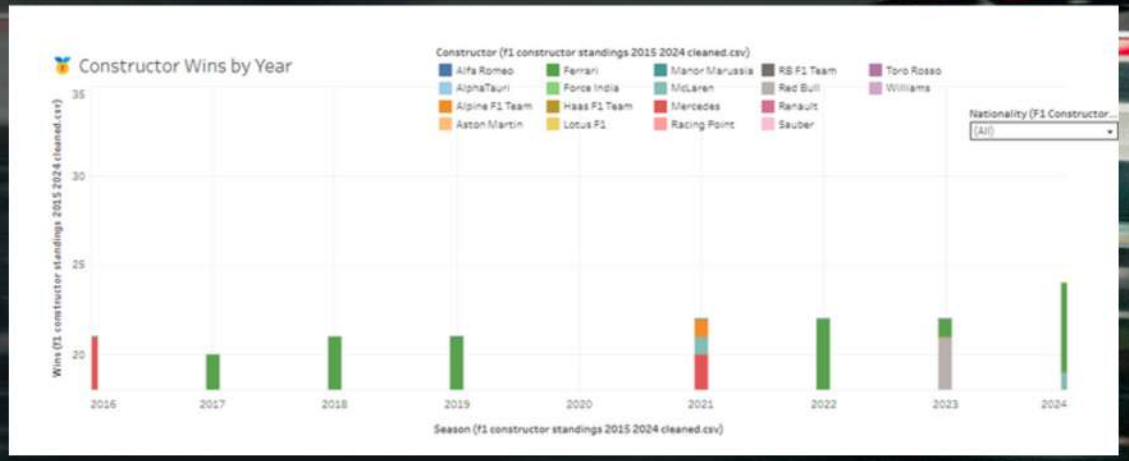


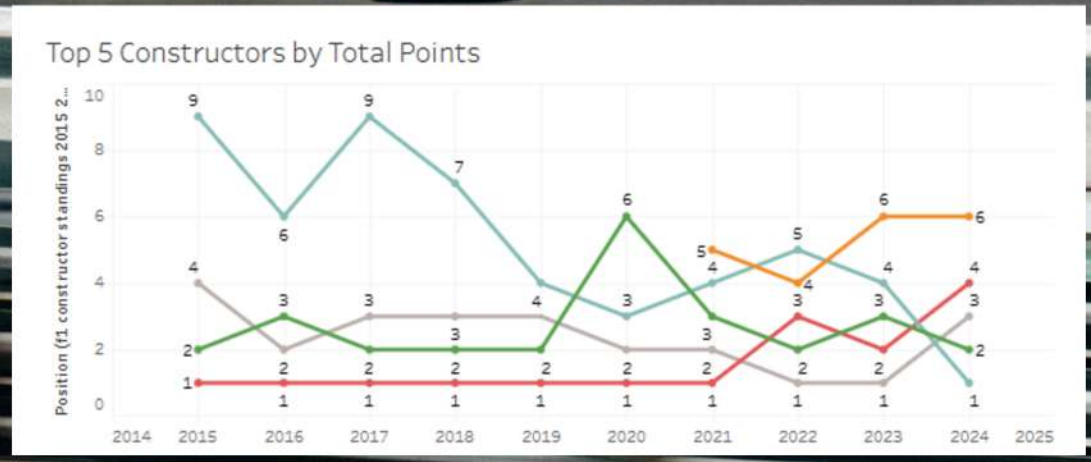
- Pitstop duration
- pitstop frequency duration

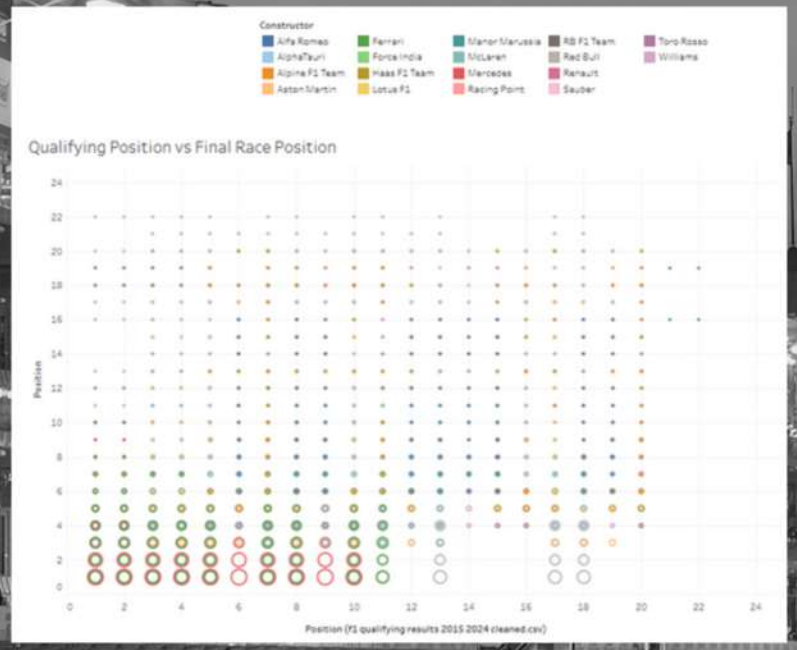
CONSTRUCTORS

- Points by constructors per year
- Constructors wins by year
- Constructors points over seas







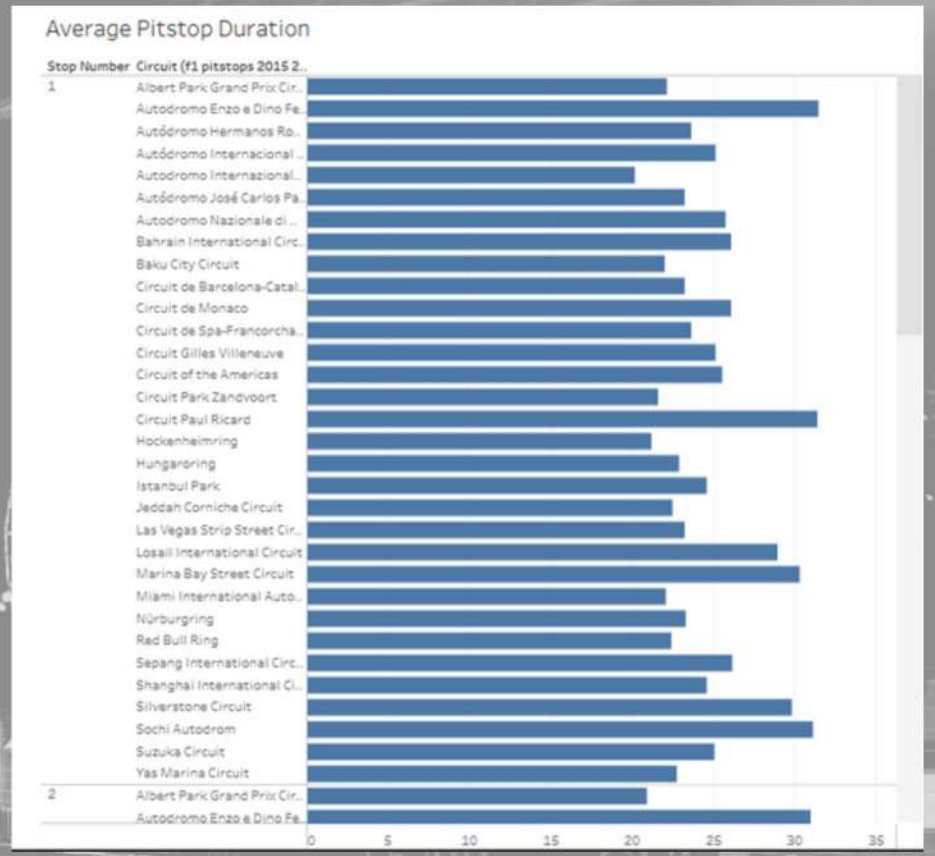


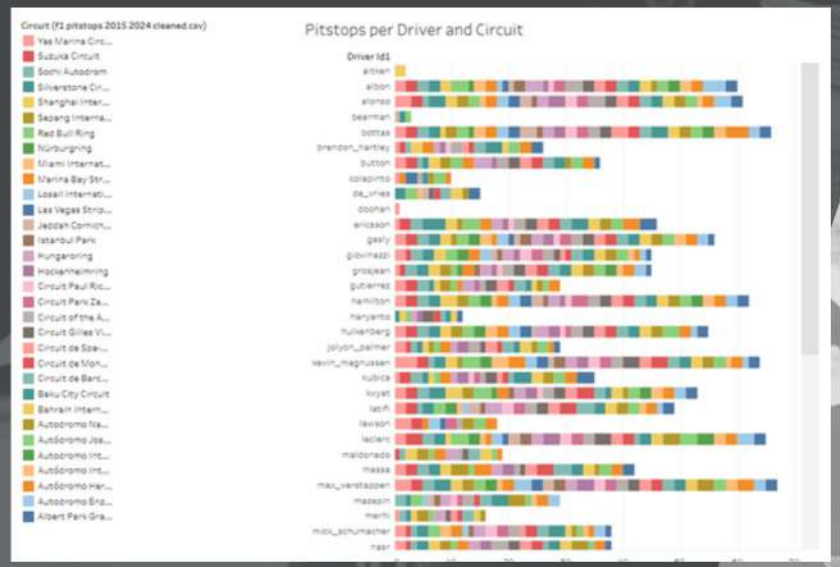
STRATEGY IN MOTION

- Pitstop per driver per race
- Qualifying position vs final race position

PITSTOP PATTERNS

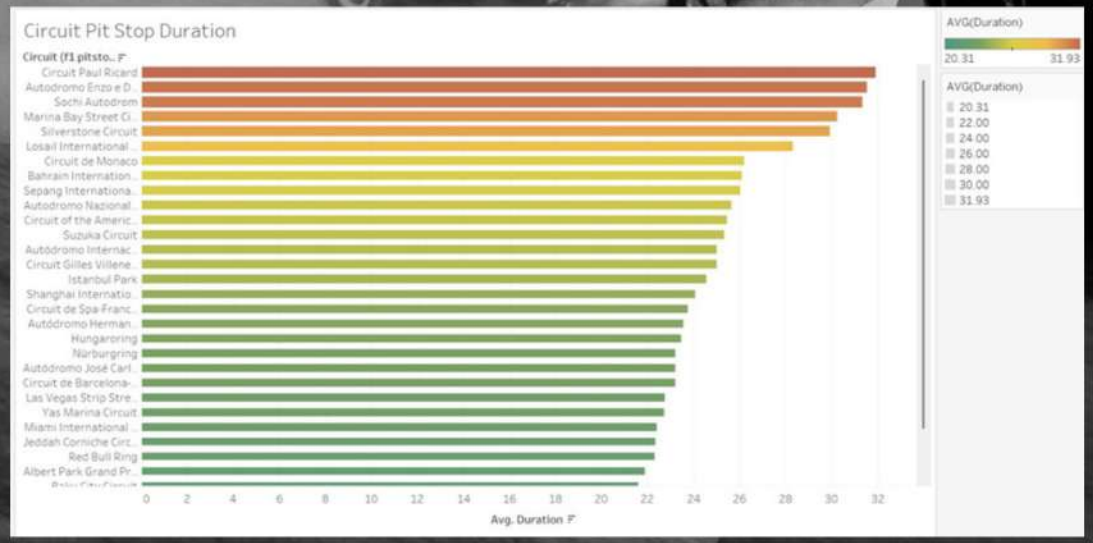
- Average Pitstop Duration
- Pitstops per Driver and Circuit





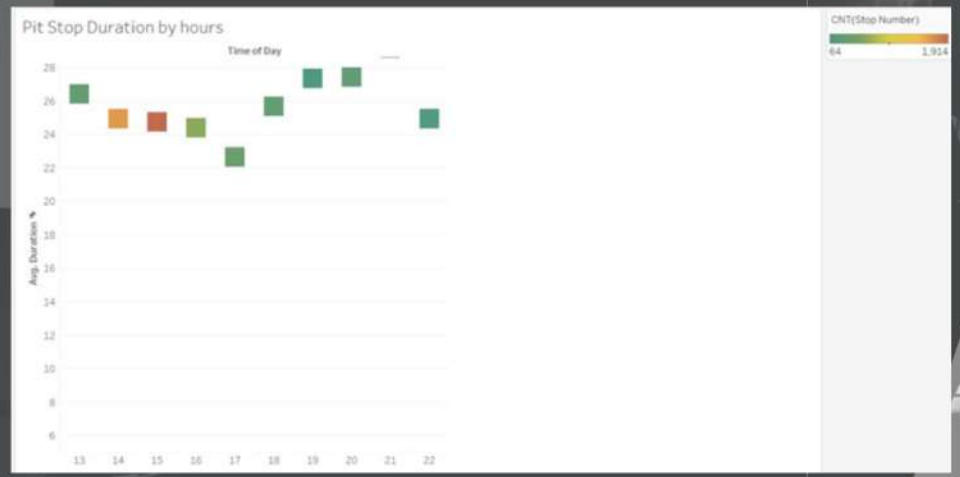
RACE LENGTH BY COUNTRY

- Longest Race Country
- Circuit Pit Stop Duration



TIMING TRENDS

- Average Pit Stop Duration per Driver
- Pit Stop Duration per hours



OUTCOMES & LEARNINGS

📈 **End-to-End Machine Learning Workflow:**

- We successfully trained and optimized a classification model to predict Formula 1 race winners, achieving strong performance metrics and validating predictions with real-world results.

🗃️ **Relational SQL Schema Implementation:**

- We designed and implemented a normalized PostgreSQL schema to organize race, driver, constructor, and standings data, simulating a production-ready data warehouse structure.

📊 **Interactive Dashboard Development:**

- Using Tableau, we created visualizations to explore pit stop durations, race lengths, and driver/constructor performance — supporting both analytical and storytelling perspectives.

🔁 **Iterative Model Optimization:**

- We documented the improvement of model accuracy through hyperparameter tuning, architecture changes, and regularization, reinforcing key machine learning concepts.

REFERENCES

- Jolpi.ca. (n.d.). F1 API – Historical Formula 1 data (Ergast). Retrieved in April 6th, 2025, from: https://api.jolpi.ca/ergast/
- Ergast Developer API. (n.d.). Motorsport Results Database. Retrieved in April 6th, 2025, from: https://ergast.com/mrd/
- PostgreSQL Global Development Group. (n.d.). PostgreSQL Database System. Retrieved in April 6th, 2025, from: https://www.postgresql.org/
- Quick Database Diagrams (QuickDBD). (n.d.). ER Diagram Tool for SQL Schema Design. Retrieved in April 6th, 2025, from: https://www.quickdatabasediagrams.com/
- Tableau Software. (n.d.). Tableau Public – Free Data Visualization Tool. Retrieved in April 6th, 2025, from: https://public.tableau.com/
- TensorFlow Developers. (n.d.). TensorFlow – Open-source Machine Learning Framework. Retrieved in April 6th, 2025, from: https://www.tensorflow.org/
- Google. (n.d.). Google Colab – Python Notebooks in the Cloud. Retrieved in April 6th, 2025, from: https://colab.research.google.com/
- The pandas Development Team. (n.d.). pandas: Data Analysis and Manipulation Tool. Retrieved in April 6th, 2025, from: https://pandas.pydata.org/
- Pedregosa, F. et al. (n.d.). scikit-learn: Machine Learning in Python. Retrieved in April 6th, 2025, from: https://scikit-learn.org/
- SQLAlchemy Project. (n.d.). SQLAlchemy – The Database Toolkit for Python. Retrieved in April 6th, 2025, from: https://www.sqlalchemy.org/
- Hunter, J. D. et al. (n.d.). Matplotlib – 2D Python Plotting Library. Retrieved in April 6th, 2025, from: https://matplotlib.org/
- GitHub, Inc. (n.d.). GitHub – Code Hosting and Version Control Platform. Retrieved in April 6th, 2025, from: https://github.com/
- Markdown Guide. (n.d.). Basic Syntax and Formatting for GitHub READMEs. Retrieved in April 6th, 2025, from: https://www.markdownguide.org/


THANK YOU FOR YOUR ATTENTION
aramco
aramco
F1
1
Mobil
BYBIT