



PREDICTING FORMULA 1 LAP TIMES

By Hina Farooq

Sector 1

Introduction

- Background on Formula 1
- Importance of track design
- The problem
- Project Objective

Sector 3

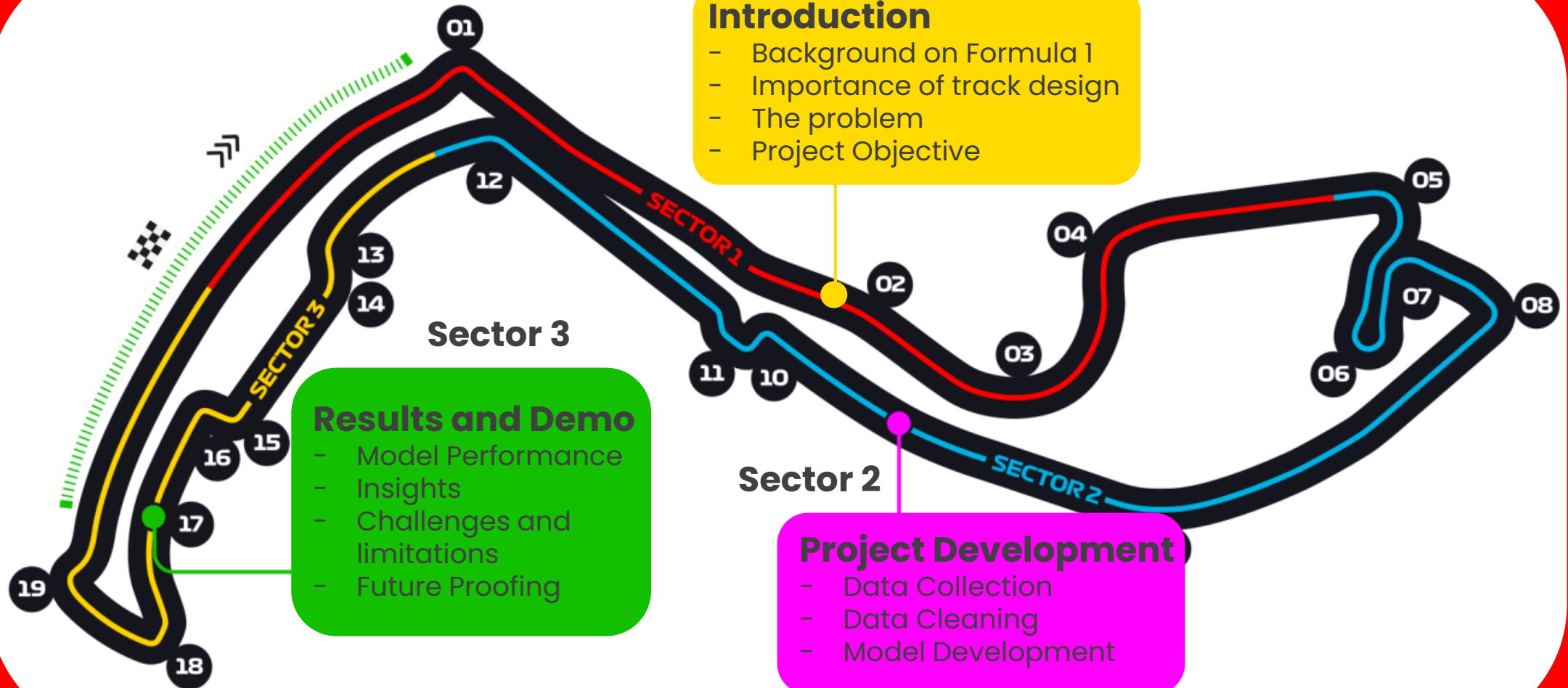
Results and Demo

- Model Performance
- Insights
- Challenges and limitations
- Future Proofing

Sector 2

Project Development

- Data Collection
- Data Cleaning
- Model Development



SECTOR 1

FORMULA 1(01)





ABOUT FORMULA 1

- Highest class of open wheeled auto racing.

Drivers

Cars

Tracks



Tracks



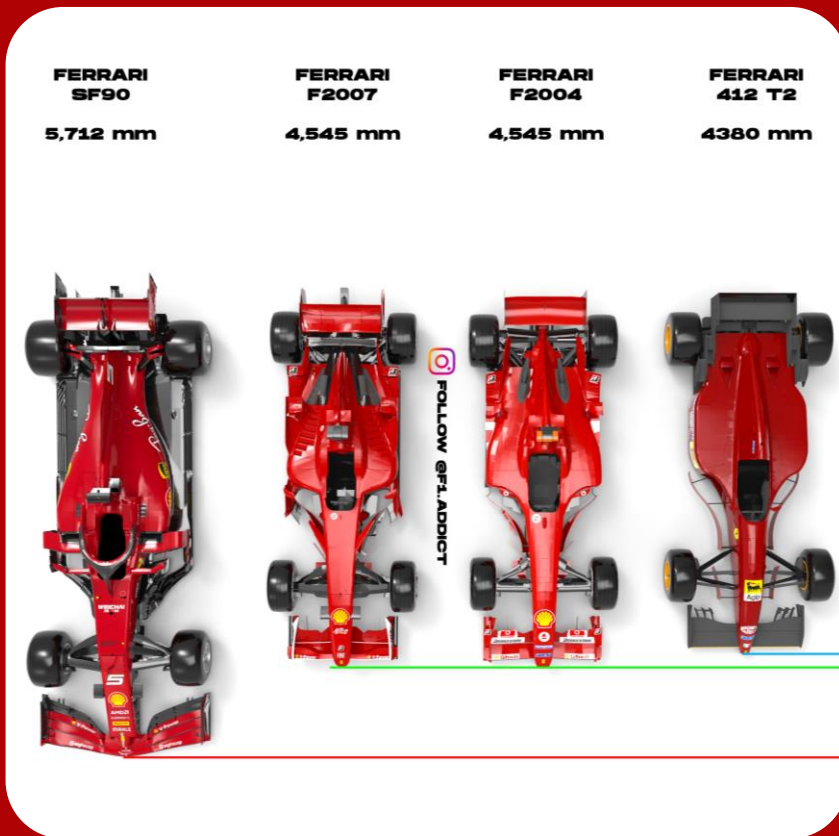
Race-Tracks

Permanent tracks built specifically for motorsport events

Street Circuits

Circuit composed of closed-off public roads





The cars have undergone significant advancements in speed and technology.

The tracks have stayed the same.

FIA has some regulations that an F1 track must meet

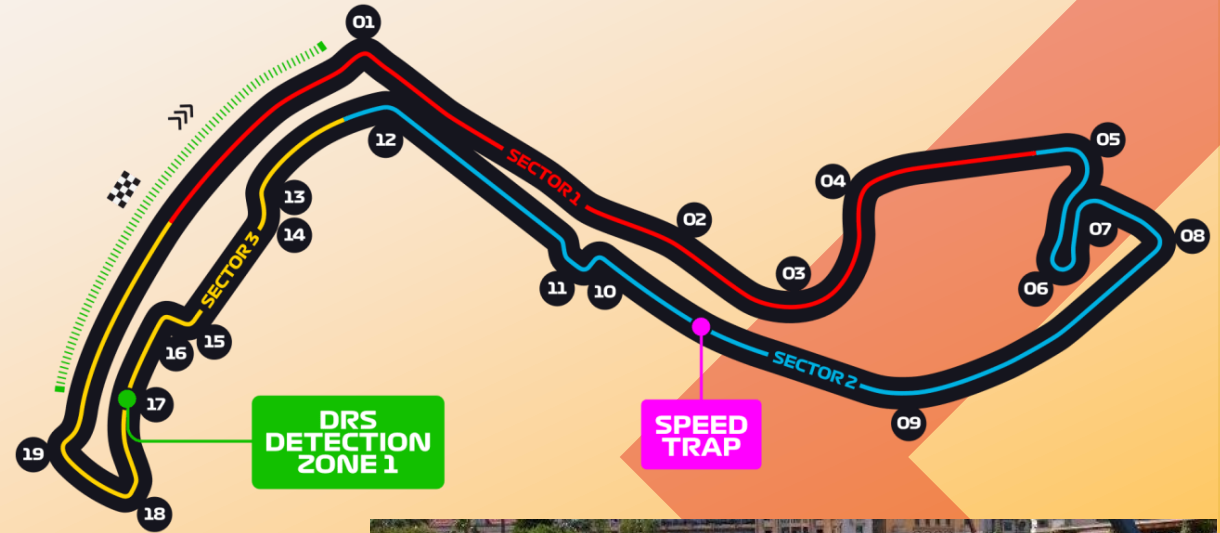


Circuit de Monaco

- Hosted 70 Grand Prix races over 75 years.
- Known for its tight corners and narrow streets.
- Narrowest Section = 7.5m

Does Monaco meet FIA standards?

NO!



Circuit de Monaco

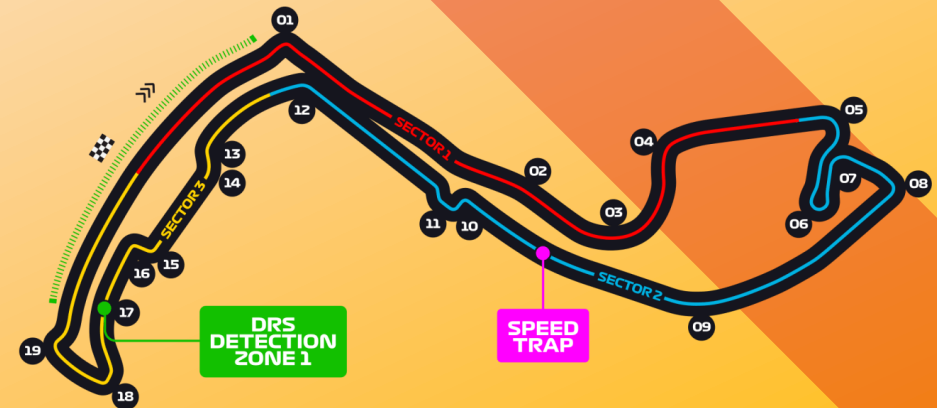
- Hosted 70 Grand Prix races over 75 years.
- Known for its tight corners and narrow streets.
- Narrowest Section = 7.5m

	FIA Regulation
Length	$3.5\text{km} < x < 7\text{km}$
Width	$12\text{ m} < x$

Why is this a problem

16 out of the last 20 winners won from pole.

UNEVENTFUL AND PREDICTABLE RACE



Predicting Lap Times From User Input



Project Scope

- Generating lap times based on user input.
- User enters in track features (length of track) → Lap Time Prediction.
- Simplified version of a track generator
- A unique solution – existing systems only offer visual representations, lacking actionable insights.

**Does something
like this already
exist?**





SECTOR 2

Project Development



3 Key Stages

Data Collection

- Web Scraping

Challenges:

- Manual imputation

Data Collected:

Fastest Lap times and track detail of 77 Circuits

Data Cleaning

- Handling missing data.
- Removing Outliers
- Standardising all data into a consistent format

Fill in null values with mean or FIA standards.

Model Development

- Feature Engineering
- Model Selection
- Performance Metrics



Data Cleaning

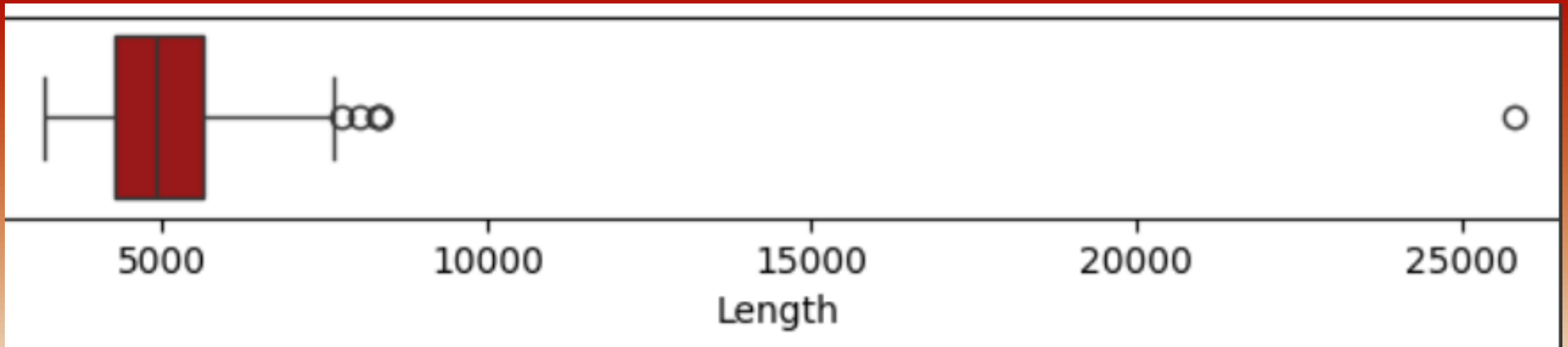
Standardising Data:

- ALL units to metres
- Lap Time converted to seconds

Handling outliers :

Pescara Circuit – length 25800 (almost 5 x longer than the average circuit)

Not representative of modern circuits



3 Key Stages

Data Collection

- Web Scraping

Challenges:

- Manual imputation

Data Collected:

Fastest Lap times and track detail of 77 Circuits

Data Cleaning

- Handling missing data.
- Removing Outliers
- Standardising all data into a consistent format

Fill in null values with mean or FIA standards.

Model Development

- Feature Engineering
- Model Selection
- Performance Metrics



Model Development

Feature Engineering:

Interaction Terms created between highly correlated features.
e.g. Race Distance x Longest Straight x Length

Model Selection and Testing

Linear Regression Models → Simple and effective for continuous predictions.

Performance Metrics:

RMSE, MAPE, R^2 ,

Final Model

Model Features:

- 1) Number of Corners
- 2) Length of Longest Straight
- 3) Length of Circuit
- 4) Type of circuit (Street or Road)



Results

SECTOR 3

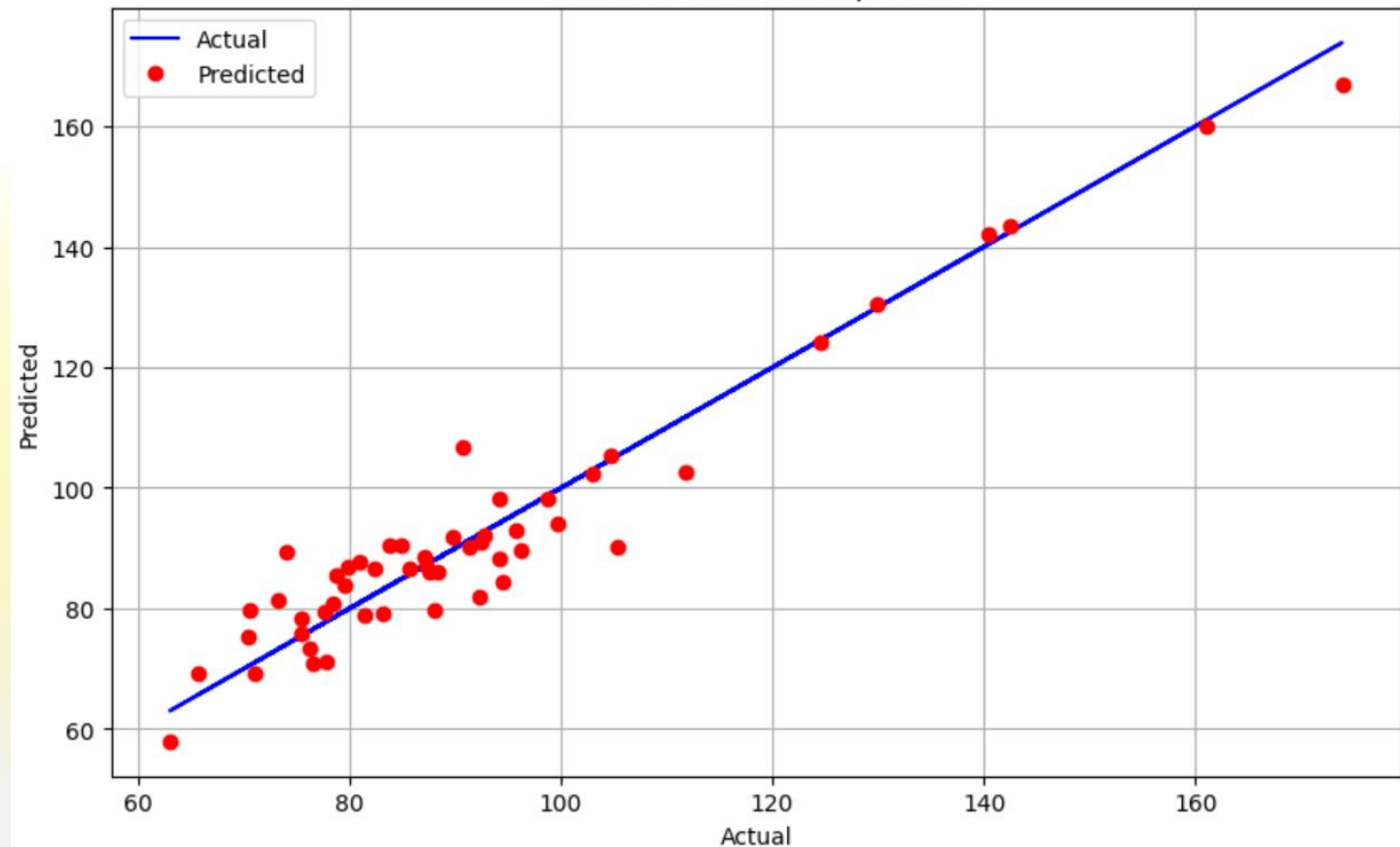


Model Performance

Predicted Values follow the same pattern as the actual values.
Note: This Graph represents the actual vs predicted lap times of Training Data.



Actual vs Predicted Lap Times



Performance Metrics

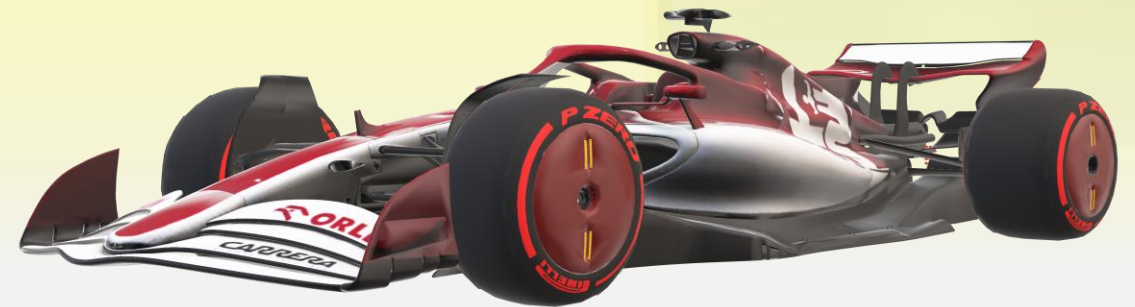
RMSE = 6.4 seconds

MAPE = 5.3%

$R^2 = 0.928$

Impact of Performance Metrics on Lap Time Predictions

- In the world of F1, 6.4 seconds is a significant margin.
- However, this model is still highly useful as it provides accurate estimations within a reasonable range.



LIMITATIONS

- Lack of available data on older circuits.
- The model doesn't account for elements like weather conditions
- Extremely small dataset

WHAT I WOULD DO DIFFERENTLY

- Collect the fastest lap time from each race over the past 20–30 years, instead of just the fastest time for each circuit.
- Include wider range of variables like track conditions, weather etc.

FUTURE PROOFING

- Continuously updating the data as new circuits emerge and new fastest laps are recorded.



Unrealised Goals

Intended Model Functionality:

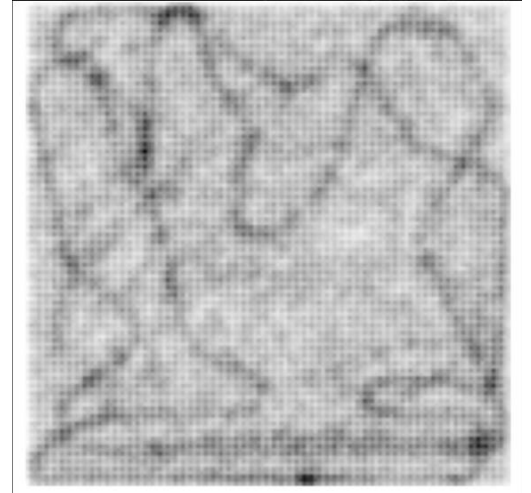
- Predict Lap Times based on user input of track features.
- Generate a visualisation of the track based on input data

Challenges:

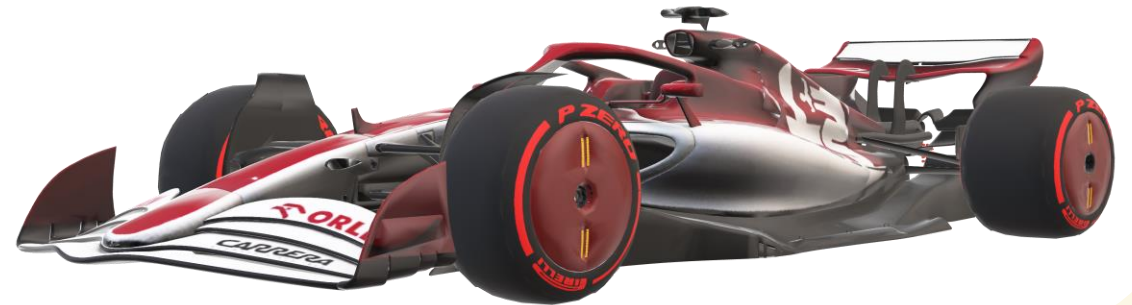
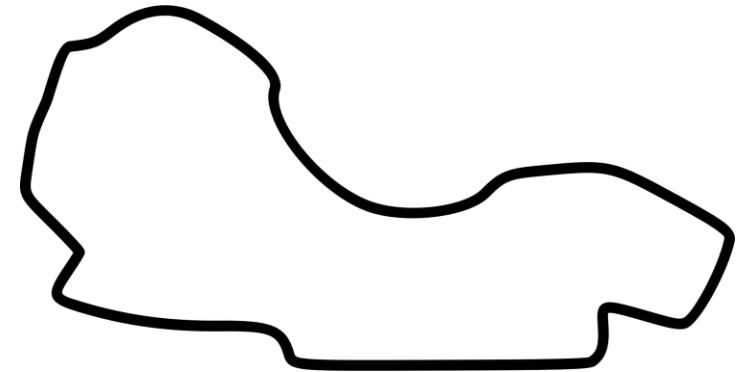
- Model was unable to integrate visualisation capabilities.
- Model underperformed even with the data it was trained on.

Combination of limited time and unforeseen technical challenges this feature was not implemented.

Track generated:



Track Data Input:



Tezos

THANK
YOU