

# Progress log

## 11-23-2024

- Home
  - Worked on project introduction
  - Draft data science questions
- Data Collection
  - Search for data sources
  - Public APIs
  - F1 Official website

## 11-24-2024

- Data Collection
  - record data
    - Race Information - (2000-2023)
    - Driver Standings
    - Circuit Info
    - Constructor Standings - not required - analysing only drivers' performance
    - weather data for each race
  - Text data
    - news api
    - news on top 10 drivers
- Data Cleaning
  - Clean Race Info: json
  - clean news data - for sentiment analysis
  - drivers standings
  - race information

## 11-25-2024

- Data Cleaning
  - include url in the race dataset
  - follow the data cleaning process on the collected data
- data collection
  - weather data
    - wiki - based on the race url from the dataset
- EDA
  - start woking on EDA
    - Text analysis
      - word clouds
      - interpretaion

– Focus on DSAN 5100 Project, 6300 Project and Final –

## **12-11-2024**

- Complete EDA
  - Word Clouds
  - Points / Driver and Constructor
  - Line plot: points / driver for different teams - different colors
  - distribution of DNF per team
  - Trend in points for constructors - for top teams only
  - pair plot for finish categories
  - work on the theme

## **13-11-2024**

- Complete Unsupervised learning
- collect data on pit stops
- clean pit stop data
- data set for clustering - race track features

## **14-11-2024**

- Complete Supervised learning
- create categorial variables for bivar. and multivar. classification

## **15-11-2024**

- Technical Pages - Topic Introductions
- Explain algorithms
- analysis
- interpretations
- customize plots - font size
- code documentation - detailed