597 DATA WRANGLING

DATA ALCHEMISTS

KAMALA PRERNA NOOKALA AND GANESH ARPAN NOOKALA

INTRODUCTION

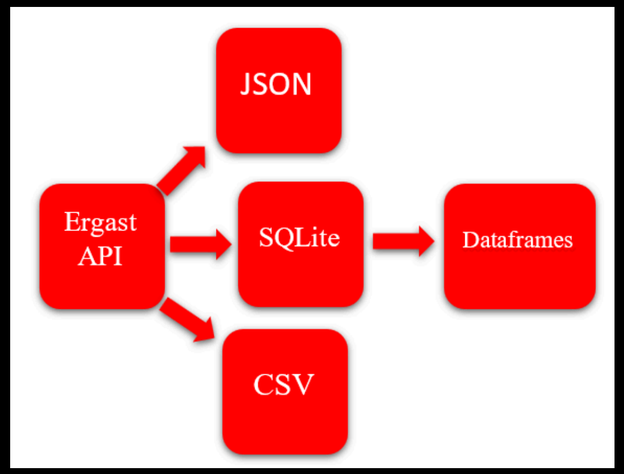
In this project, we delve into decades of Formula 1 data to craft detailed visualizations and predictive models that illuminate the sport's strategic depth. Through rigorous analysis of race outcomes, driver performance metrics, and team histories, we aim to expose patterns that offer new insights into the tactical nuances of Formula 1.

DATA RETRIEVAL

We employ the Ergast API, a comprehensive source for historical Formula 1 data, to fuel our analyses and visualizations.

RAW TO TABULAR DATA

The data extracted from the API consists of deeply nested JSON, which we process, normalize, and extract to obtain only the necessary data. This data is then loaded into a SQLite database, simultaneously stored in cache, and also saved in CSV format. Subsequently, we create dataframes for visualization and modeling.

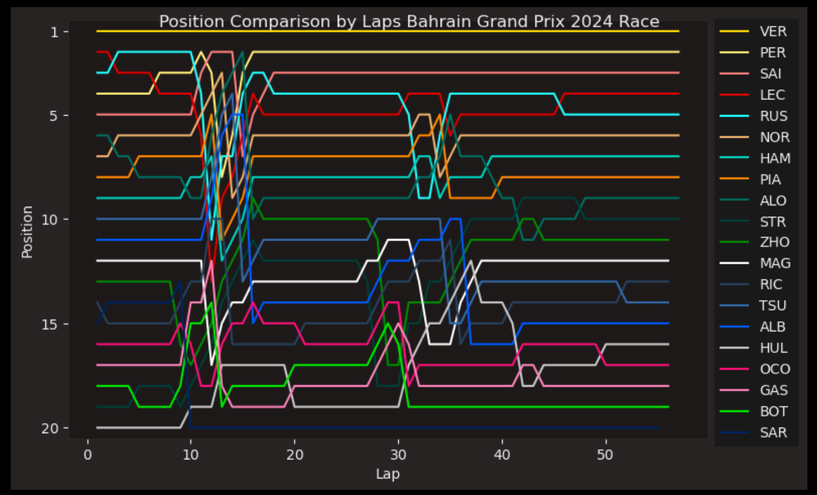


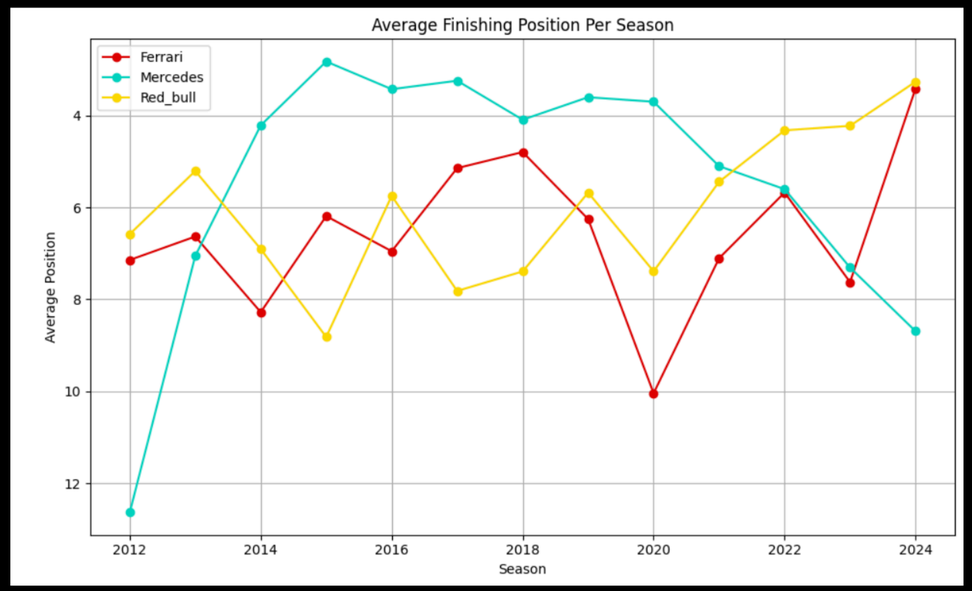
DATA ENRICHMENT

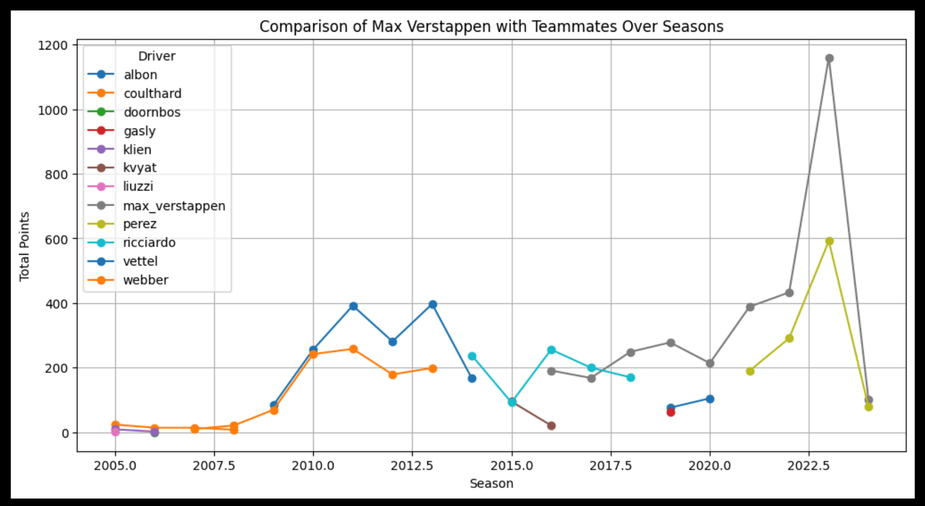
We've computed essential features within our dataset for each season, including:

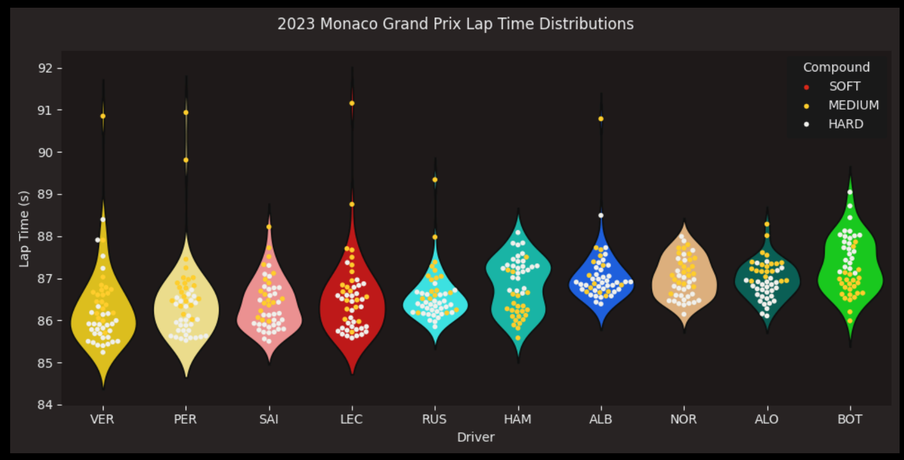
1. Average Finishing Position
2. Podium Finishes
3. Average Pit Stop Count
4. Average Pit Stop Duration
5. Total Points

VISUALIZATIONS









COMMIT GRAPH

Graph	Description	Date	Author	Commit
	gn178commits origin **Named plots accurately**	28 Apr ...	arpannooka...	2b748a8c
	origin/HEAD origin/main Merge branch 'main' of https://github.com/stats-at-Rutgers/group-project-data_alche...	28 Apr ...	premanook...	530ef1a9
	Updated data loading procedure	28 Apr ...	premanook...	85d96b91
	gh-pages origin Built site for gh-pages	28 Apr ...	arpannooka...	fff685f5
	Merge pull request #2 from stats-at-Rutgers/gn178commits	28 Apr ...	arpannooka...	ea4a429c
	Added work on data analysis	28 Apr ...	arpannooka...	621a6122
	Cleaned qmd files, and added necessary information	28 Apr ...	arpannooka...	1c79da70
	main Created sqlite database, and streamlined data extraction	26 Apr ...	premanook...	bd5be1dc
	Built site for gh-pages	11 Apr ...	premanook...	26280b85
	Update Automated Retrieval, Raw to Tab, Data Cleaning, Summary Statistics, Data Visualization	11 Apr ...	premanook...	1f2acab1
	Built site for gh-pages	11 Apr ...	premanook...	db0fd435
	Update Summary Statistics	11 Apr ...	premanook...	556e7394
	Built site for gh-pages	10 Apr ...	premanook...	2c57ef44
	Update Data Retrieval, Wrangling and Cleaning	10 Apr ...	premanook...	f3c1d64f
	Update Data Sources, and Enrichment sections	10 Apr ...	arpannooka...	3c5749cf
	Setting up GitHub Classroom Feedback	10 Apr ...	github-class...	9b04a1f8
	origin/feedback GitHub Classroom Feedback	10 Apr ...	github-class...	32549298
	Initialize gh-pages	10 Apr ...	github-class...	efe0dd0c
	Initial commit	10 Apr ...	github-class...	708d4fc7

CIRCUIT MAPS