UNIVERSITI TENAGA NASIONAL

PUTRAJAYA CAMPUS

BACHELOR OF INFORMATION TECHNOLOGY (INFORMATION SYSTEM) (HONS.)

SEMESTER 1 2O22/2023

CGMB4113:

Information Visualization

CAN SCUDERIA FERRARI WIN THE 2022 FIA FORMULA 1 CONSTRUCTOR CHAMPIONSHIP?

Muhammad Afiq Bin Khushairi
IS0105573

Contents

Contextual Planning

Introduction

Formula One is the highest class of international open wheel car racing. A formula One season consists of series of races known as Grand Prix which been held worldwide on multiple circuits. A points system is used at Grands Prix to determine two annual World Championships: one for drivers, the other for constructors. Each driver must hold a valid "Super License" which is the highest class of racing license issued by the FIA. All the races must be run on the tracks graded "1". One of the oldest team in the sport is Scuderia Ferrari and arguably one of the best Formula 1 constructor team of all time. The team had 16 world constructor title and 15 world driver championship title. The team also had construct quite a number of world class drivers such as Michael Schumacher, Nikki Lauda and Sebastian Vettel. Nevertheless, the last 15 years Ferrari have struggle to keep up with the other competitors to win the world constructor and driver championship title. The 2022 FIA Formula 1 might give the team a new hope in grabing the world driver and constructor title since a major regulation update had been made by the Fédération Internationale de l'Automobile (FIA) to make the sport more competitive.

This project will portray all the insight related to the performance of Scuderia Ferrari from 2019 to mid-season of 2022 and make prediction whether if the team can win the FIA Formula 1 World Constructor Championship. The data will consist of how many points collected by the team, finished race position and car/engine reliability on each race weekend.

Objectives

1. To predict whether the Scuderia Ferrari capable of winning the 2022 Formula 1 world championship.
2. To identify Scuderia Ferrari performance from 2019 to mid-season of 2022

Messages and Insights

1. Shows the insights whether Team Scuderia Ferrari will have a chance to win the constructor championship or not in 2022 season.
2. The visualisation contains the analysis for the constructor, driver and circuit that shows the strength, the weakness and Ferrari's capabilities on the track.

Target Audience

The target audience of this data visualisation will be the motorsport (Formula 1) fans that trying to see if their favourite or rival team (Scuderia Ferrari) can win the 2022 Formula 1 Constructor Championship. Since the audience will the public viewers, the design of the graph and charts will be as minimalistic and simple as possible to grasps the audience attention and get to comprehend the insight try to shown in the visualisation.

Publishing Strategy

The visualisation will be present in a digital format in order to make the information accessible to anyone who own electronic devices such as laptops/computers, tablets and smart phones. The visualisation output will be present in two format which is a presentation deck and an infographic. The visualisation output type will be static data presentation. The graphs and charts will be produced in Microsoft Power BI.

Data Preparation

Data Source

Link: https://www.kaggle.com/datasets/rohanrao/formula-1-world-championship-1950-2020?datasetId%5C=

The dataset utilised in this project is acquired from a publicly accessible dataset titled "Formula 1 World Championship (1950 - 2022)", which was obtained from the Kaggle website for the data science community. There are 110 columns and almost 20,000 rows in the dataset. The data set being obtain form Kaggle consist of multiple data set which being divided by certain parts. There are total of 17 datasets available to be used in this project which had been download in compressed zip file from Kaggle. Some of dataset available such as “drivers”, “constructors” and “results”. For the purpose of this project, the dataset will be use only 8/6 out of 17 datasets available.

Data Preparation Activities

Throughout this stage of the process, the raw dataset that was collected will be processed into a finished dataset that can be applied to the visualisation. In order to make this data transformation, the dataset will be going through cleaning process by removing the unwanted data in the in the dataset. There were several tools available to assist in the data cleansing process. The tool that will be using in this project will be Python Jupyter Notebook. The dataset consists of all data regarding Formula One races from 1950 until 2022.

The dataset will be load into python to start the data cleaning process. All the dataset involve will be removes the unwanted attributes and rows to only keep the data that related, in this case will be the data from 2019 till mid-season of 2022 that had been collected.

```
In [17]: import pandas as pd
         import numpy as np
```

```
In [18]: df=pd.read_csv("drivers.csv")
         df
```

Out[18]:

	driverId	driverRef	number	code	forename	surname	dob	nationality	url
0	1	hamilton	44	HAM	Lewis	Hamilton	1985-01-07	British	http://en.wikipedia.org/wiki/Lewis_Hamilton
1	2	heidfeld	\N	HEI	Nick	Heidfeld	1977-05-10	German	http://en.wikipedia.org/wiki/Nick_Heidfeld
2	3	rosberg	6	ROS	Nico	Rosberg	1985-06-27	German	http://en.wikipedia.org/wiki/Nico_Rosberg
3	4	alonso	14	ALO	Fernando	Alonso	1981-07-29	Spanish	http://en.wikipedia.org/wiki/Fernando_Alonso
4	5	kovalainen	\N	KOV	Heikki	Kovalainen	1981-10-19	Finnish	http://en.wikipedia.org/wiki/Heikki_Kovalainen
...	...	...	...	...	...	...	...	...	...
849	851	aitken	89	AIT	Jack	Aitken	1995-09-23	British	http://en.wikipedia.org/wiki/Jack_Aitken
850	852	tsunoda	22	TSU	Yuki	Tsunoda	2000-05-11	Japanese	http://en.wikipedia.org/wiki/Yuki_Tsunoda
851	853	mazepin	9	MAZ	Nikita	Mazepin	1999-03-02	Russian	http://en.wikipedia.org/wiki/Nikita_Mazepin
852	854	mick_schumacher	47	MSC	Mick	Schumacher	1999-03-22	German	http://en.wikipedia.org/wiki/Mick_Schumacher
853	855	zhou	24	ZHO	Guanyu	Zhou	1999-05-30	Chinese	http://en.wikipedia.org/wiki/Guanyu_Zhou

```
In [23]: df.drop(["driverRef","url"], axis=1, inplace=True)
         df
```

Out[23]:

	driverId	number	code	forename	surname	dob	nationality
0	1	44	HAM	Lewis	Hamilton	1985-01-07	British
1	4	14	ALO	Fernando	Alonso	1981-07-29	Spanish
2	8	7	RAI	Kimi	Räikkönen	1979-10-17	Finnish
3	9	88	KUB	Robert	Kubica	1984-12-07	Polish
4	20	5	VET	Sebastian	Vettel	1987-07-03	German
5	154	8	GRO	Romain	Grosjean	1986-04-17	French
6	842	10	GAS	Pierre	Gasly	1996-02-07	French
7	807	27	HUL	Nico	Hülkenberg	1987-08-19	German
8	815	11	PER	Sergio	Pérez	1990-01-26	Mexican
9	817	3	RIC	Daniel	Ricciardo	1989-07-01	Australian
10	822	77	BOT	Valtteri	Bottas	1989-08-28	Finnish
11	825	20	MAG	Kevin	Magnussen	1992-10-05	Danish

Solution Development

Visual Mapping & Design Decision Explanation

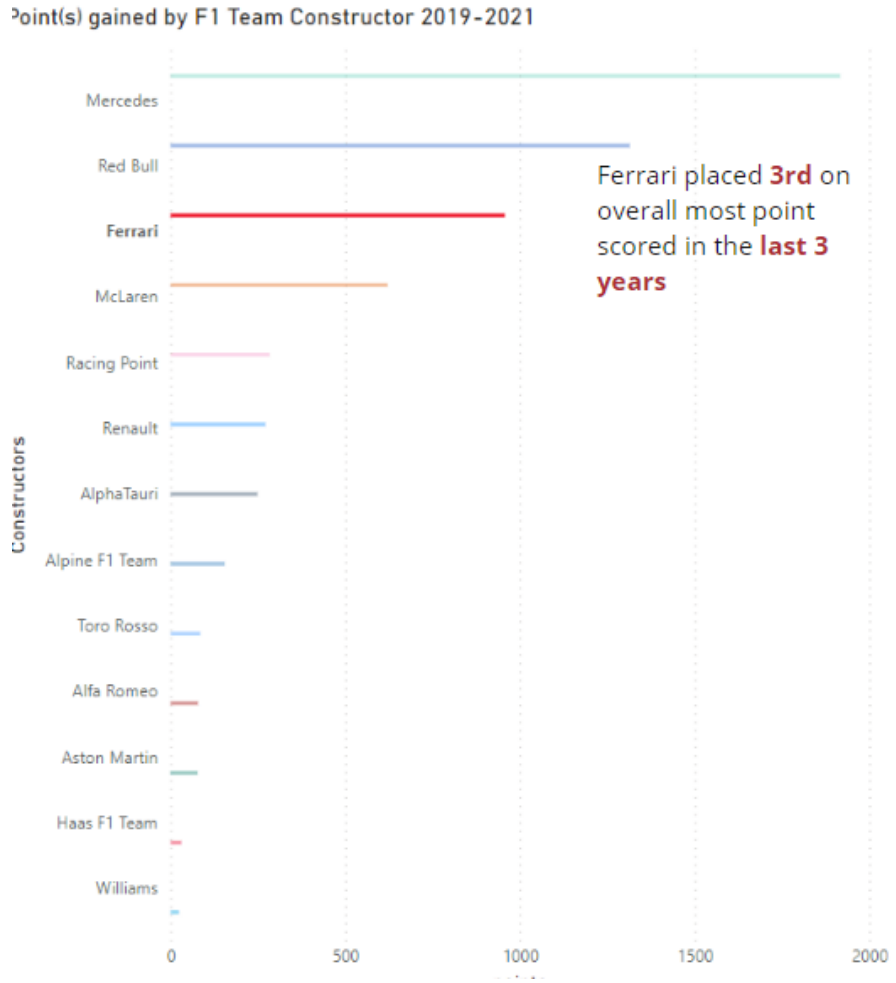


Figure above shown points being scored by each team that compete in year 2019 to 2021. Ferrari team placed in third as most points scored team overall for the last three years. This type of data suitable to be visualised in bar chart to display data and use to compare each category values. Using bar chart for this data also can show each data frequency evenly. As shown in the figure, audience can see clearly that Mercedes leading in term of overall points scored from 2019 to 2021. It also will be easier to read as the total of category is 10 and the name of the category will be visible.

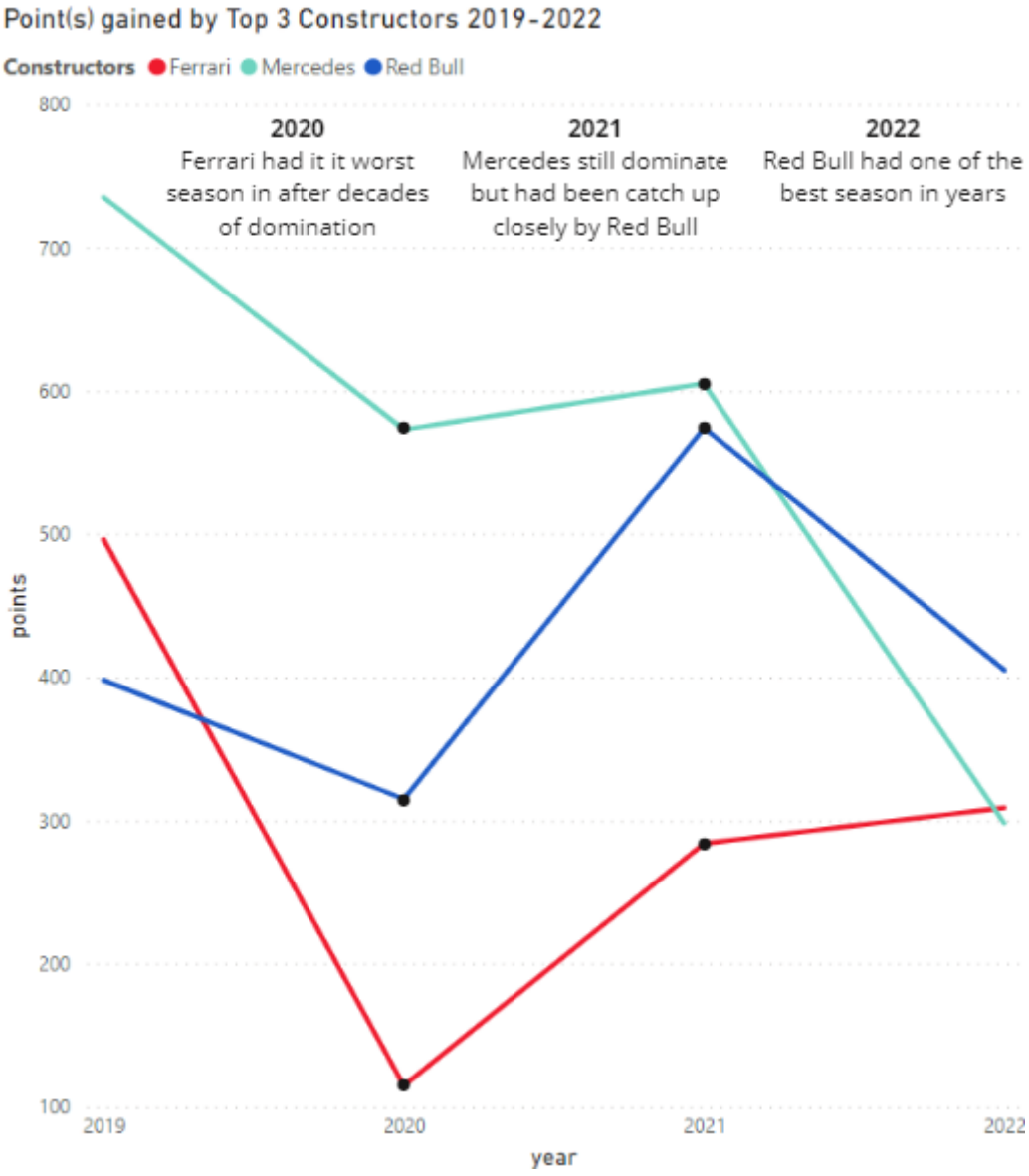


Figure above shown points being scored by each top 3 constructors from 2019 to mid-season 2022. This data being visualise as line chart to show the trend of each top 3 constructor performance each year. Each colour for the line represents their team official colour. Line graphs can be used to compare changes for diverse categories over the same time period. As in figure, the line shown trend of performance by the constructor each year based on total point scored.

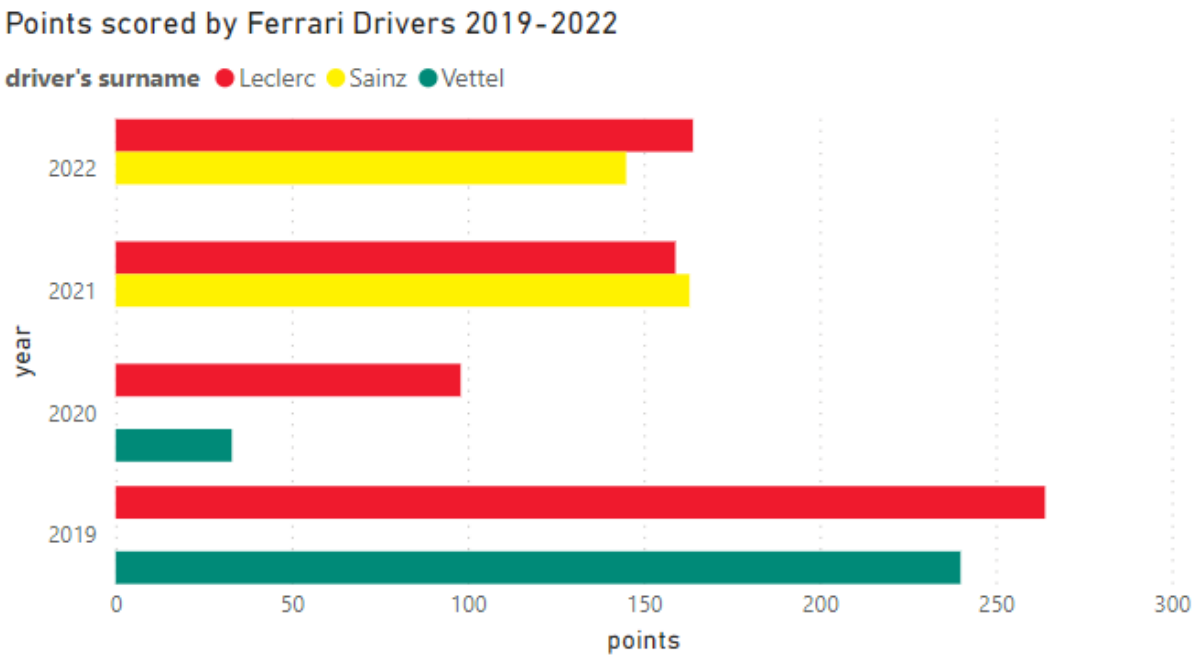
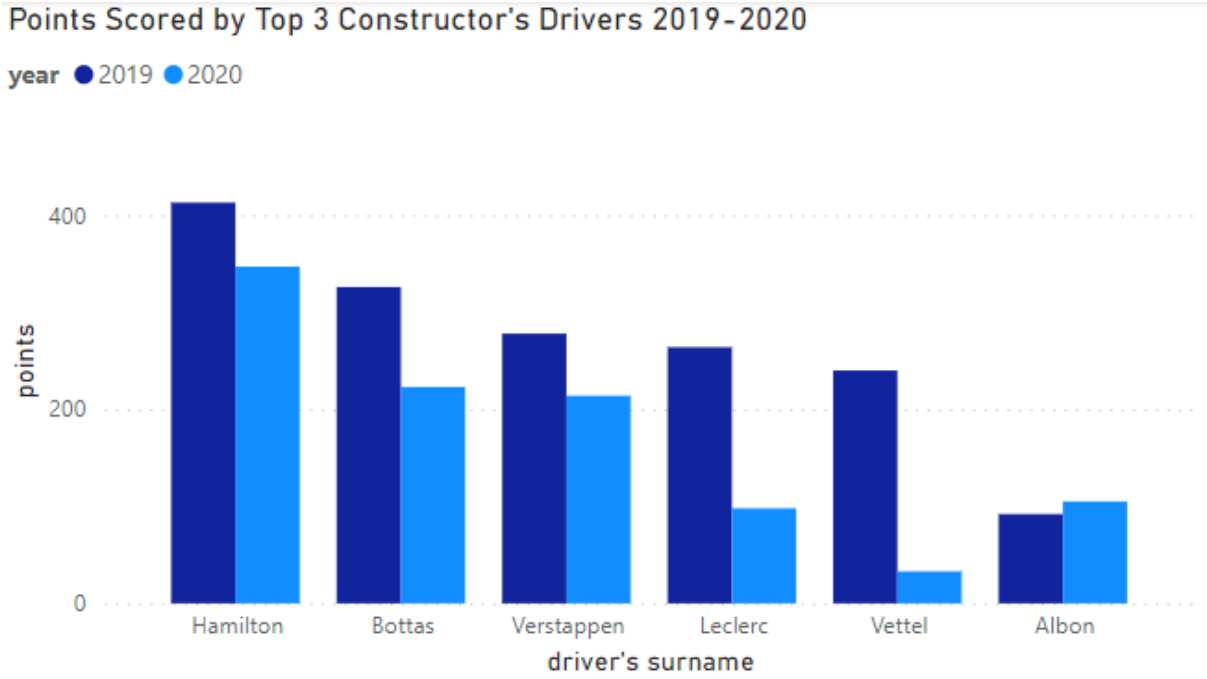


Figure above shown points scored by the Ferrari drivers from 2019 to mid-season 2022 in a bar chart. This chart being chosen to visualise the performance of Ferrari drivers based on point collected each year. For the year 2019 and 2020, the driver for Ferrari is Vettel and Leclerc while in 2021 and 2020 is Leclerc and Sainz. Points shown in figure above symbolise the performance by drivers each year.



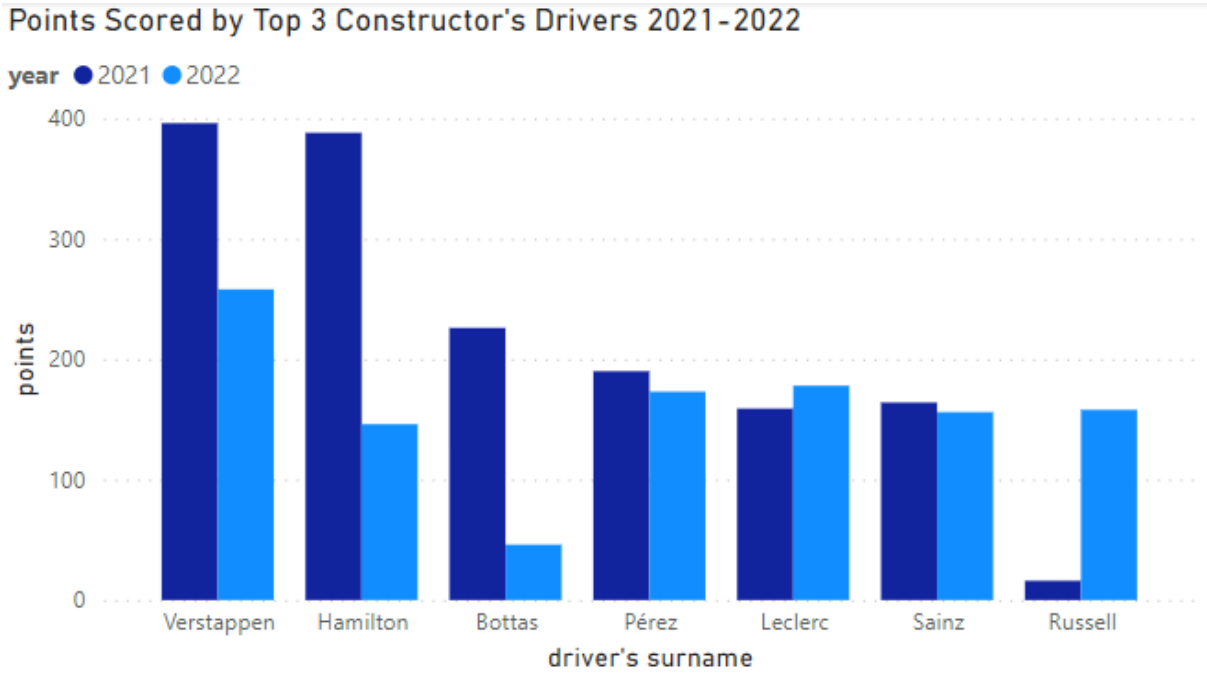


Figure above shown points scored by the Top 3 constructor's drivers from 2019 through the middle of the 2022 season. The appropriate way to display this kind of data is as a bar chart, which allows one to analyse the number of times each driver earned points each year.

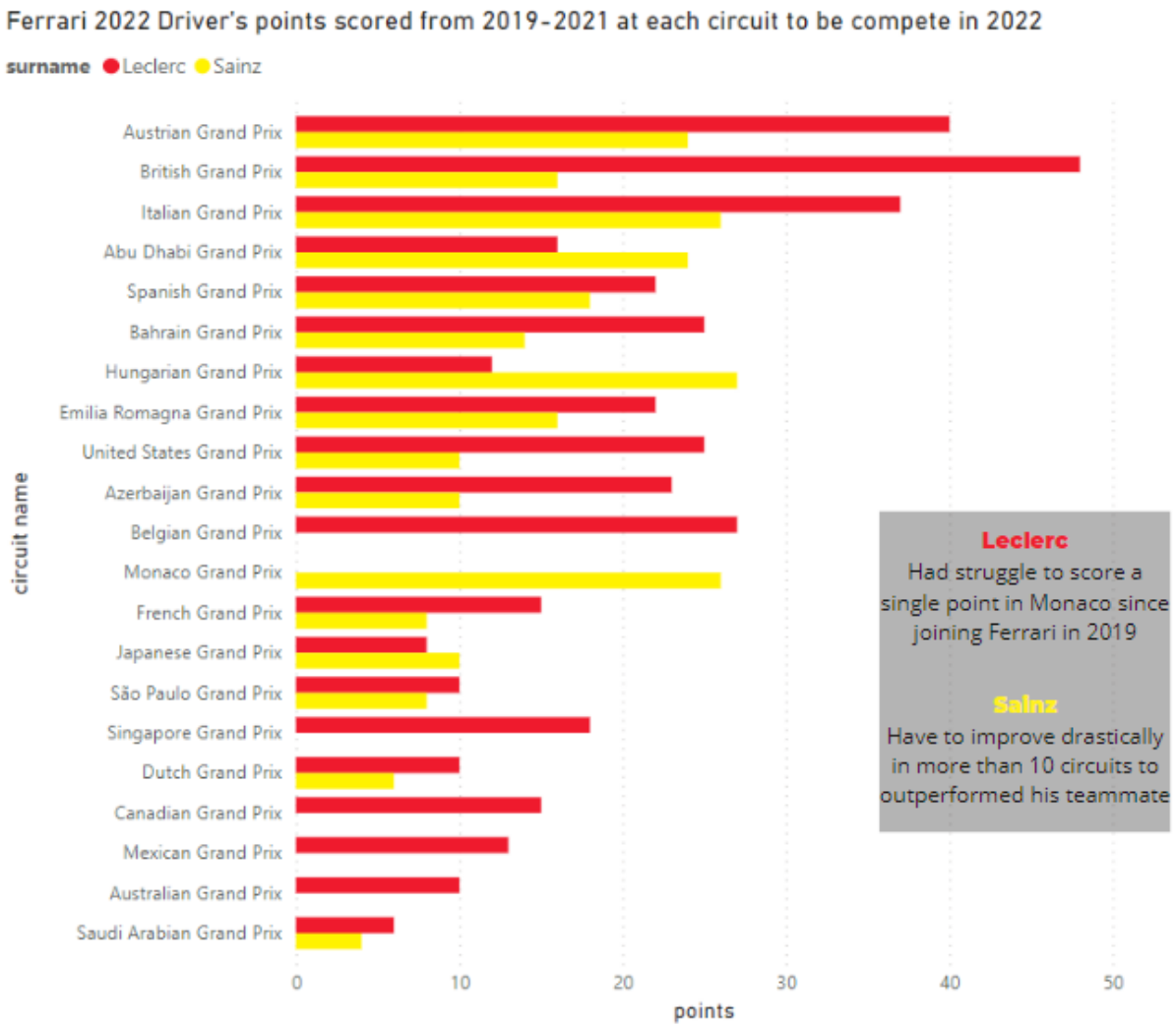


Figure above being visualise as a bar chart to represent all previous performance from 2019-2021 by each current Ferrari drivers at circuit that will be compete in 2022. This data visualised in bar chart to make all the circuit name visible and to see clearly in which circuit that the driver most performed based on the points scored at each circuit.

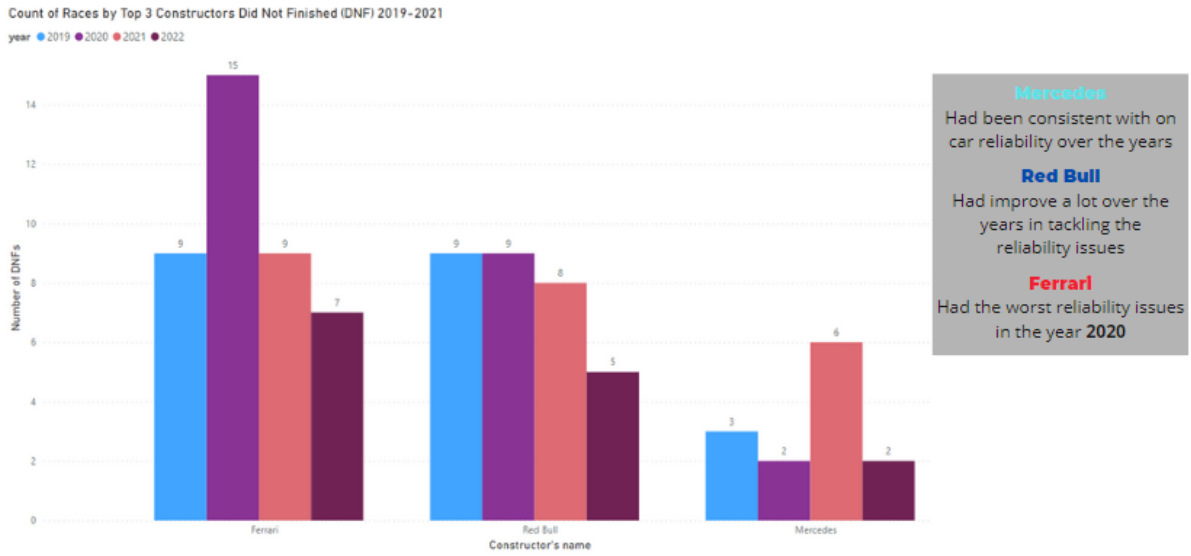
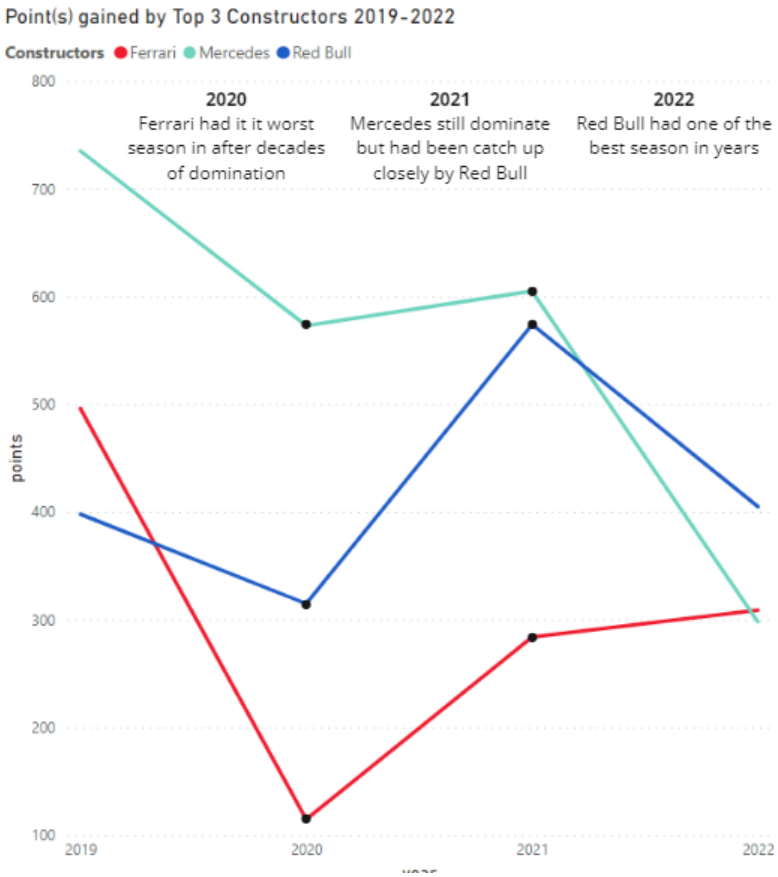


Figure above shows the number of race retirement by each top 3 constructor from 2019-2021. This column chart being chose to visualise the frequency of car retirement each year from these top 3 constructors.

Design Decision for Overall

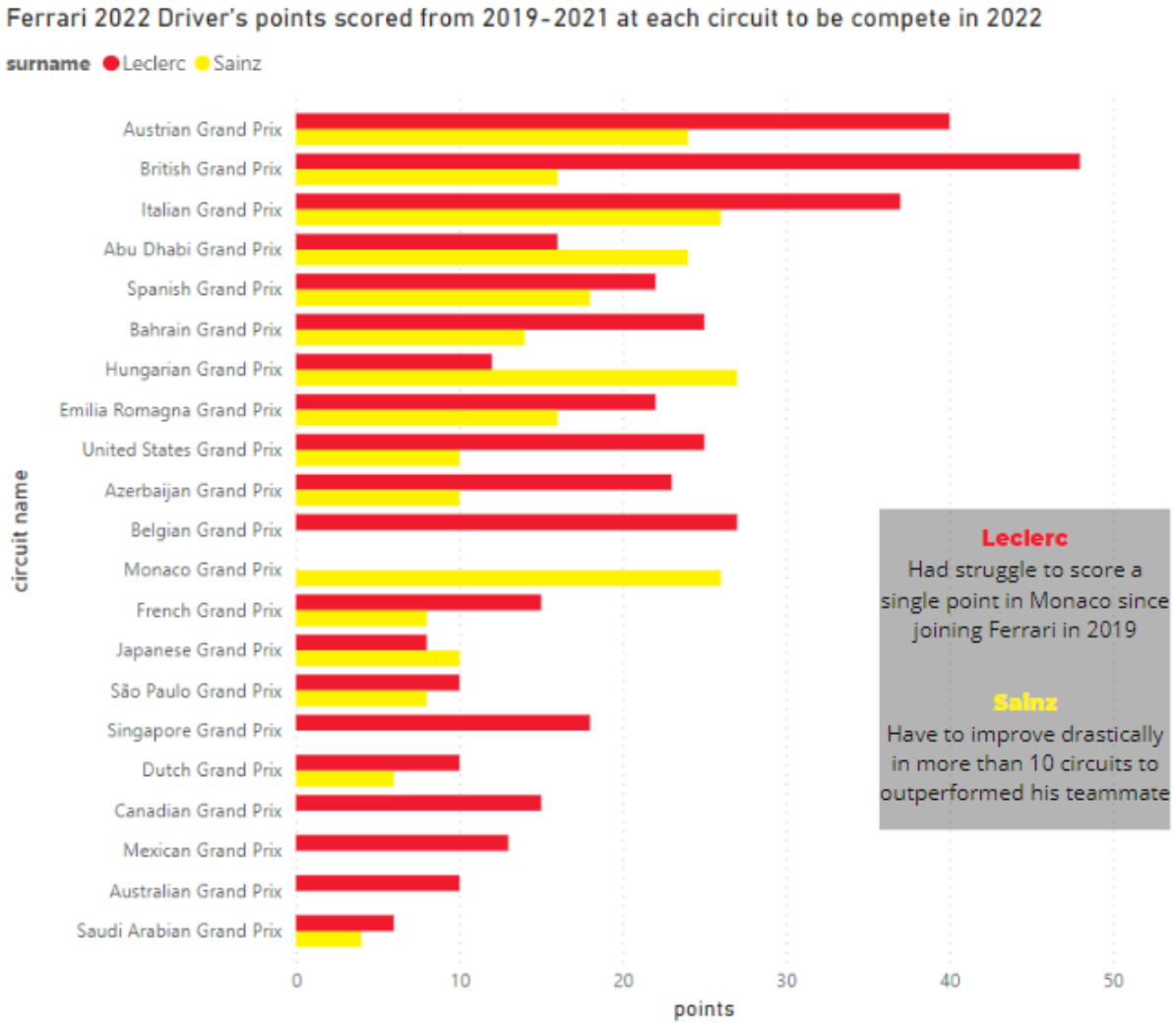
Presentation deck

1. Annotation



The annotation being added in the graph to explain on what is happening in the chart presentation.

2. Callout



Callout also being added to the charts to explain a little bit about the data that had been visualised.

3. Statement title

For all the charts in this presentation, the type of title being used is statement title as it used to describe each graph in the presentation deck.

Infographic

1. Colour scheme

   The colour scheme used for the info graphic only 4 colour schemes. This decision being made to make sure the infographic does not look cluttered by the audience and viewers.

2. Visual Hierarchy

   In order to build an infographic with visual balance and maintain the overall design's coherence, each element's arrangement and order are organised according to relevance.

Reference

- https://en.wikipedia.org/wiki/Formula_One
- https://nces.ed.gov/nceskids/help/user_guide/graph/whentouse.asp#:~:text=Line%20graphs%20are%20used%20to,for%20more%20than%20one%20group.

Presentation Video

- https://www.youtube.com/watch?v=bymyuXbkmz0

Presentation Deck

- https://www.canva.com/design/DAFVMBh0ciM/EfSxq0D17Vh5O9YzhDGKbg/view?utm_content=DAFVMBh0ciM&utm_campaign=designshare&utm_medium=link2&utm_source=sharebutton

Infographic

- https://www.canva.com/design/DAFVNEDcDOQ/KHBYGJu_ADf06sRacnBLQQ/view?utm_content=DAFVNEDcDOQ&utm_campaign=designshare&utm_medium=link2&utm_source=sharebutton