ROCKET LEAGUE

WORLD
CHAMPIONSHIP
2021 - 2022
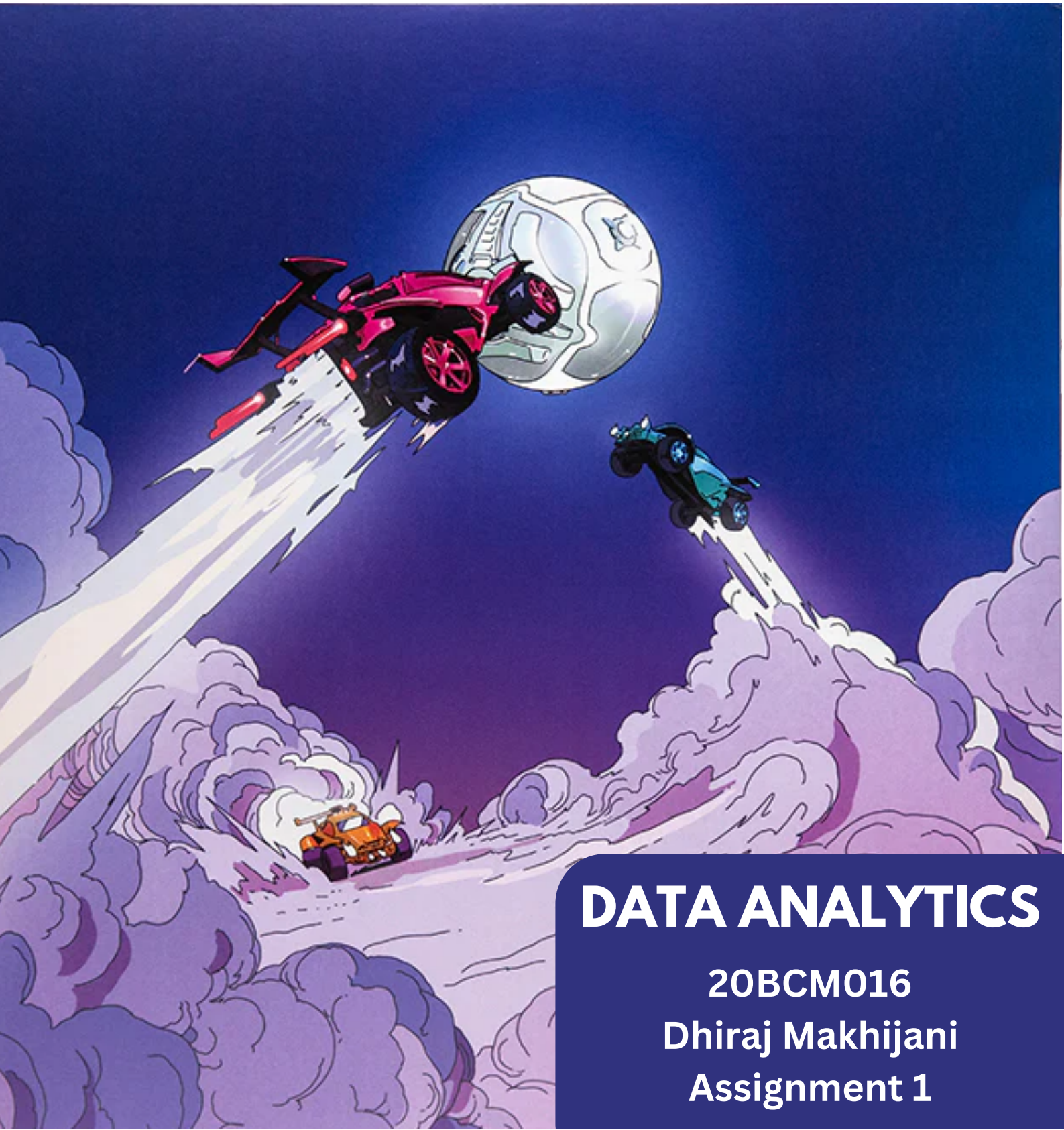
DATA ANALYTICS

20BCM016
Dhiraj Makhijani
Assignment 1

INDEX

Introduction

The Rocket League Championship Series (RLCS) is an annual Rocket League Esports tournament series produced by Psyonix, the game's developer for the game Rocket League.

The game rocket league is similar to the sport football with the added benefit that the player is able to control a flying car. The game is played in a 3v3 format with the total game length being 5 minutes albeit it can go to overtime if scores are equal and the first team to score wins the game.

In regards to rocket league gameplay, there are 6 cars, one ball, and 2 goalposts on the field. Each player controls his own respective car.
Boost pads are scattered around the field that allows the player to gain speed and fly through the air as per the amount of boost he has.

When a player reaches max speed or supersonic speed he can perform demolitions on opponents temporarily taking the opponent out of the game (About 5 seconds).

Content

RLCS allows players from all over the world to compete over a huge prize pool.

The RLCS 2021-2022 consists of a qualification split in North America, South America, Europe, Oceania, Middle East/North Africa, Asia, and Sub-Saharan Africa, and culminates in a playoff bracket with teams from those regions. The qualification rounds are played as an online round-robin tournament and the finals are played live in different cities.

3 splits - Fall, Winter, Spring - each have 3 regional events leading into an international major. Within all these events, each team collects points that will determine whether they will qualify for the World Championship or World Championship Wildcard.

Depending on the stage of an event, an RLCS match can be played to the best of 5 or 7 games, with special treatment for the Fall Split finals played in 2 winning sets of BO7.

The following events took place in RLCS 2021-2022

Fall Split
- Regionals 1, 2, 3 (All regions - Invitational Qualifiers, Closed Qualifiers and Main Events)
- Fall Major - Asia-Pacific Qualifier
- Fall Major - North America Tiebreaker (Complexity Gaming vs. Spacestation Gaming)
- Fall Major - Main Event

Winter Split
- Regionals 1, 2, 3 (All regions - Closed Qualifiers and Main Events)
- Winter Major - Asia-Pacific Qualifier
- Winter Major - Main Event

Spring Split
- Regional 1, 2, 3 (All regions - Closed Qualifiers and Main Events)
- Spring Major - Asia-Pacific Qualifier
- Spring Major - Main Event

World Championship
- Wildcard Stage (Play-In)
- Group Stage
- Playoffs

Data Summarization

The dataset **Rocket League Championship Series 2021-2022** was picked up from Kaggle, and it contains 6 CSV files. The primary CSV file chosen is **matches_by_teams.csv** and it focuses on performance given by teams on RLCS 2021-2022.

The dataset contains **10594** rows and **54** columns. Every single matchup between 2 teams is listed here.

Among 54 columns, the ones that need to be focused on as a prerequisite are described below

- **match_id** primary key assigned to Matchup between 2 teams that will be played in either best-of-5 or best-of-7
- **team_id** primary key assigned to the team
- **team_name** Name of the team that participated.
- **team_region** The region the team belongs to.
- **score** How many games did a team win in a best-of-5 or best-of-7 series?
- **winner** Was the team winner of the matchup?

The rest of the columns are **gameplay-based** columns like goals, assists, shots, saves, boost used, demos,time flying,offensive time,defensive time,etc.

We become acquainted with the following data

- **647** teams
- **9** Regions
- **5297** Matches

It is to be kept in mind that these values are still susceptible to null values.

We have **47** columns having **~0.2%** of null values. A keen observation here is that the **match_id, score, and winner,** columns do **not** have any missing values at all so we always have the outcome of a match.

Keeping that in mind, we come to the conclusion that if a match is missing **gameplay-based** values, that particular match will be missing **all of its gameplay-based** values.

team_name	team_region	core_shots	core_goals	core_saves	core_assists	...	positioning_time_neutral_third
R!OT GAMING	Oceania	NaN	NaN	NaN	NaN	...	NaN
PANIK GAMING	Oceania	NaN	NaN	NaN	NaN	...	NaN
JOE	Oceania	NaN	NaN	NaN	NaN	...	NaN
THE GROGANS	Oceania	NaN	NaN	NaN	NaN	...	NaN
RANCID RATS REBORN	Oceania	NaN	NaN	NaN	NaN	...	NaN

Grouping by region tells us that about **~600** matches was played by every region during this season.
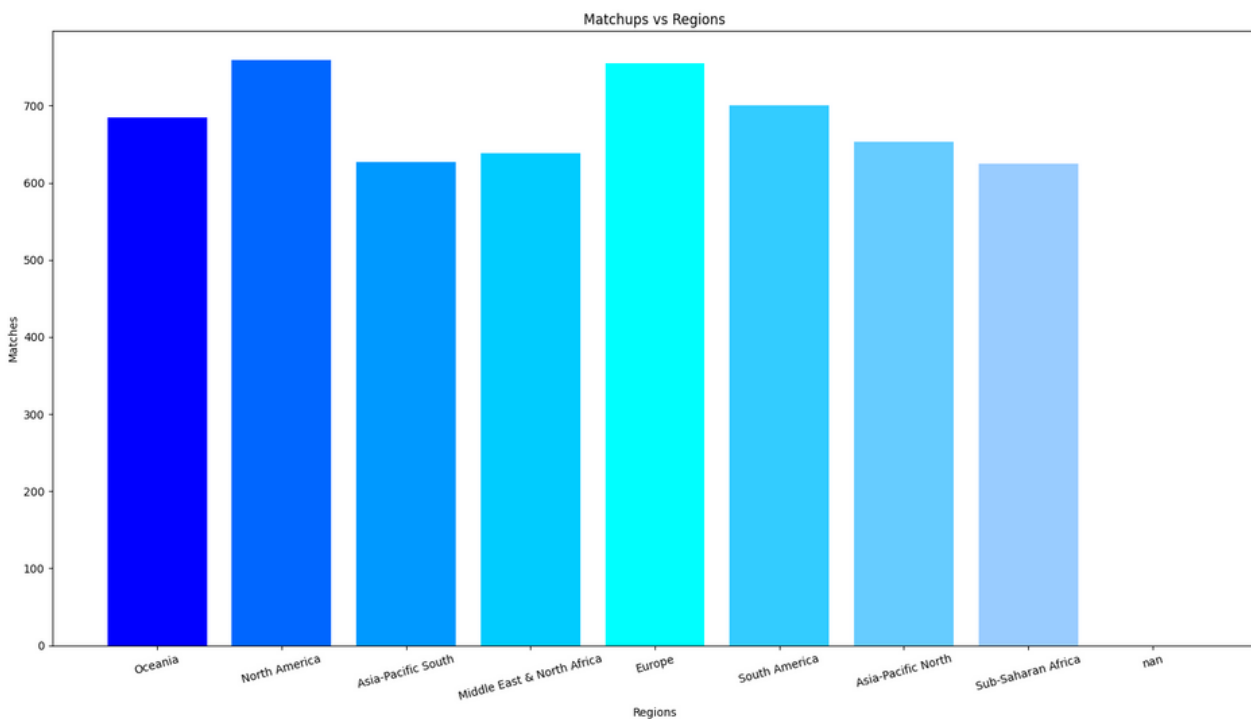
This dataset provides adequate data for us to analyze how each region performed.

We however can't know **when** or under **what** event the match was played. To handle this we can refer to another **main.csv** file which specifies each match with the event it associates with.

Objective

With the dataset in hand, we can proceed to data mine the following things-

- **Winning Team Analysis**

  Team BDS, a European team is the one that wins this RLCS 2021-2022. Studying this team's journey from the first fall split all the way to the world finals, how close it was to losing, gameplay stats and so on.

- **Ranking Points Collection**

  Within all these events, each team collects points that will determine whether it qualifies for the world championship or not. With our dataset, knowing how a team performed during a major event can let us deduce ranking point for a particular team.

- **Meta Analysis**

  With over 8 regions, it would be interesting to glance over gameplay statistics. Does the Europe region perform more demolitions than North America? Do Asia players play more defensively? These can be answered through our dataset.

- **Predictions**

    Having complete data on how much a team performed can be used to make predictions on how the team will perform in RLCS 2022-2023.

- **Ranking Points Collection**

    Within all these events, each team collects points that will determine whether it qualifies for the world championship or not. With our dataset, knowing how a team performed during a major event can let us deduce ranking point for a particular team.

- **Visualization**

    Comparing among different teams,regions can be done by plotting multiple graphs.

Preprocessing

Starting with **Data Cleaning**, as discussed before teams have no gameplay data at all that cannot be dropped or replaced with mean values.

Gameplay data like goals, assists and saves can occur only in real-time during a match. It cannot be replaced on the basis of mean or any other imputation. The matchups could have goals 10-5 or 3-1, it is not feasible to **make up those gameplay-based values**.

These rows can also not be **dropped**, we know the **outcome** of the match, every team is of significance here.

Some teams do have **team region** missing, this is corrected as a team shows up in multiple matches thereby its team region can be fetched from a team's other matches.

No outlier removal is required, the gameplay data is collected via an API which works upon Rocket League's replay system. Every data collected represents everything that was done in the match. In terms of Team Performance Analysis, every Team was given an even playing field amongst everyone else.

For **Data Transformation**, we look at specifically gameplay columns numbered from 12 to 50. By observing the minimum value and maximum value of each column it is observed that the difference exceeds 1000 to about 10000 for each column out there. Hence, minMax algorithm is applied. Z-score algorithm isn't required because we don't need negative values.

As described in Data Summarization, in order to know which **event** a match took place under, we have to refer to another file. This is where we will integrate **main.csv** with our main file. The **main.csv** contains both **match_id** and **event_id,event_name.** We are able to **map, event_name** with the respective **match_id** in our main file.

This allows us to cover **data discretization** as data can now be divided on per event basis too.

Data Reduction is achieved via **attribute subset selection.** Data can be split on the basis of
- Teams
- Region
- Events

We have reached the stage wherever we are able to perform various data mining-esque operations.

Europe: Teams vs Goals
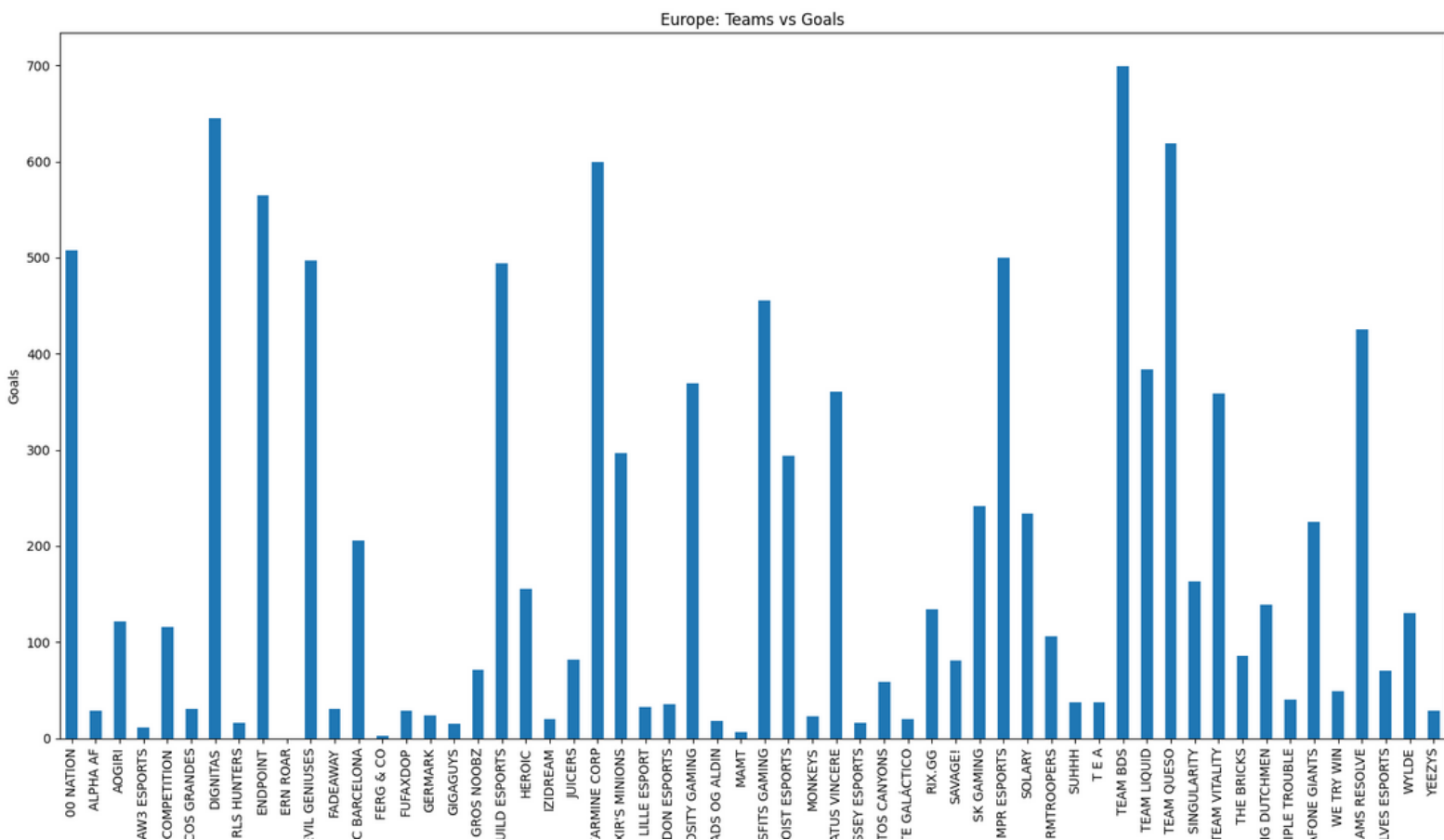
Outcome

The data preprocessing applied to our dataset makes our dataset feasible for data mining techniques,implementations now.