CS 5010: Semester Project Proposal

Manpreet Dhindsa, Robel Semunegus, Amber Curran

mkd8bb, rzs5yf, akc6be

Group 5

**INTRODUCTION**

Our purpose of analyzing this data set is to detect trends and patterns of the NBA over the last 20 years. We want to understand how teams and players have changed over time and how that may impact future seasons.

**THE DATA**

Our data set will come from the NBA Team Stats from espn.com for seasons 2001-2020. This data set will include statistics about each team in the NBA as well as the statistics for each player on each team. We chose this data set because we wanted to understand how teams and players performed over time. In addition, working with sports data allows us to derive interesting trends that we can compare to a real world application. We plan to webscrape espn.com in order to obtain this data, so we will not need to carry out any data cleaning since it is already in a clean and organized format. If there are any outliers or missing information, we will perform the necessary data cleaning.

**EXPERIMENTAL DESIGN**

We will use the data from espn.com and use web scraping to obtain the necessary data. Then, we will organize the data in order to derive different types of analysis. We expect to manipulate data frames and create visualizations to support analyses and trends.

**PROJECT MANAGEMENT**

We plan to have weekly meetings to serve as check-ins and to delegate tasks to each other. As we break down the project further, roles will be created in order to order to efficiently complete the project.

**RESULTS**

We plan to employ user input in order to allow users to interact with the data and produce the results they are interested in. We also will allow them to print out visualizations of the trends, as well. Through our report and video presentation, we will outline in greater detail our analysis and conclusions.

**TESTING**

We will conduct appropriate unit testing in order to confirm our code works to our expectations.

**OUTCOME**

This could be used for future seasons and could be a way to predict the potential teams and players based on the trends and patterns from the historical data.