Report Title

Student Name 1   
Institutional ID:   
Email:

Student Name 2   
Institutional ID:   
Email:

Student Name 4   
Institutional ID:   
Email:

Student Name 3   
Institutional ID:   
Email:

*Abstract*— A general small description of your project. What did you want to do? What data did you use? How? What did you find?

# Introduction

Movies are an important piece of culture in our society. The best films are remembered for generations and can make an unknown actor’s career, while bad movies can have the opposite effect. The revenue aspect is also crucial, since production companies will reward an excellent film with sequels and a well-performing actor with more castings. Our analysis focuses exactly on this: we examine a movie dataset to learn about patterns and what makes a film a “good” one.

The movies have been analysed through three datasets: the most comprehensive one includes metadata such as the title, the genre, and the average vote given by reviewers relative to circa 45000 movies; the second dataset is a reduced version of the former, with only a third of the observations, and provides the movie plots; the last dataset has been collected from the *Academy Awards Database* and provides a list of all the Oscar Award winners from 1927 till 2020 with relevant metadata.

# DATA

## Dataset 1 - Movie metadata

The main dataset. It includes information about movies, with the earliest being released in 1874/12/09 and the latest planned for 2020/12/16, which are collected in the following format:

* An index relating to the collection the movie belongs to, if it exists;
* The movie budget;
* The genres the movie is categorised as;
* The original language;
* The original title;
* The international title;
* An overview, which briefly describes the movie;
* The popularity score, which takes into account the page views on the IMDB website;
* A list of the production companies involved in making the movie;
* A list of the production countries;
* The date of release;
* The revenue;
* The runtime in minutes;
* The languages spoken in the movie;
* The status of the movie, such as “released”, “post-production” or “planned”;
* A tagline;
* The average of the votes given by reviewers;
* The number of votes that have been given to the movie by reviewers.

## Dataset 2

Describe your dataset here….

# Exploratory Analysis

A preliminary data exploration to derive hypotheses….

## Hypothesis 1: “TITLE”

This is about….

## Hypothesis : “TITLE”

This is about…

# Methdology

What metrics are you planning to use? What methodology?

## Metric 1: “TITLE”

This is related to hypothesis…

## Metric 2: “TITLE”

This is related to hypothesis…

# Analysis Design

Here you have to talk about all the choices done to cope with your Big Data scenario.

How did you plan solving scalability issues?

## Database

Adopted storage solution and how did you use it.

## Dimensioning the Problem

Make some considerations here about how a possible hardware architecture should look like in a “standard” scenario (the one you imagine to be possibly dealing with). How many nodes? How would you configure OpenStack?

**IMPORTANT: This section reports only assumptions, you DO NOT have to build a real cluster**. **You are going to use Hadoop and Spark in SINGLE NODE mode (ON YOUR LOCAL LAPTOP/PC - no real Big Data Problem).**

However, you still need to think about scalability and try to design your overall solution.

In your project you must use **ONLY** your PC and VirtualBox. As seen during the practical lessons, **all the technologies adopted (Hadoop, Spark, etc.) must be installed and configured in your local machine (in a Virtualbox virtual machine).**

## Hadoop/Hive

Describe here your experience using Hadoop/Hive in your project. What did you use it for?

## Spark

Describe here your experience using Spark in your project. What did you use it for? Which modules? Why?

# Analysis description and Results

This section is devoted to present your analysis, your plots, your findings and any limitation.

Provide here all technical details of interest (how did you address specific problems and why).

Also try to “tell” your story…an analysis is always a “tell” of something leading to a finding. Try to explain what led you to your choices and how did you manage to obtain your findings.

##### Conclusion

Provide here a summary of your exploitation. Explain what did you find in summary.

##### References

1. Here any reference you want to incude…also links or documents available online…papers you found, etc.