9/1/2022

Proposal

Video game sales

Ebowusim Michael
STUDENT NUMBER - 4050630

Table of Content

Introduction

I have decided to take on a project for the open programme. The following document contains information about the project, stating my understanding of the data, the goal of the project as well as graphical representations of information I found after exploring the dataset.

Also give some insight on how the project might impact society in both positive and negative ways. This document will also include the proposal in regards to achieving the project goal as well as the constraints and risks I may face that could impact the quality of the project.

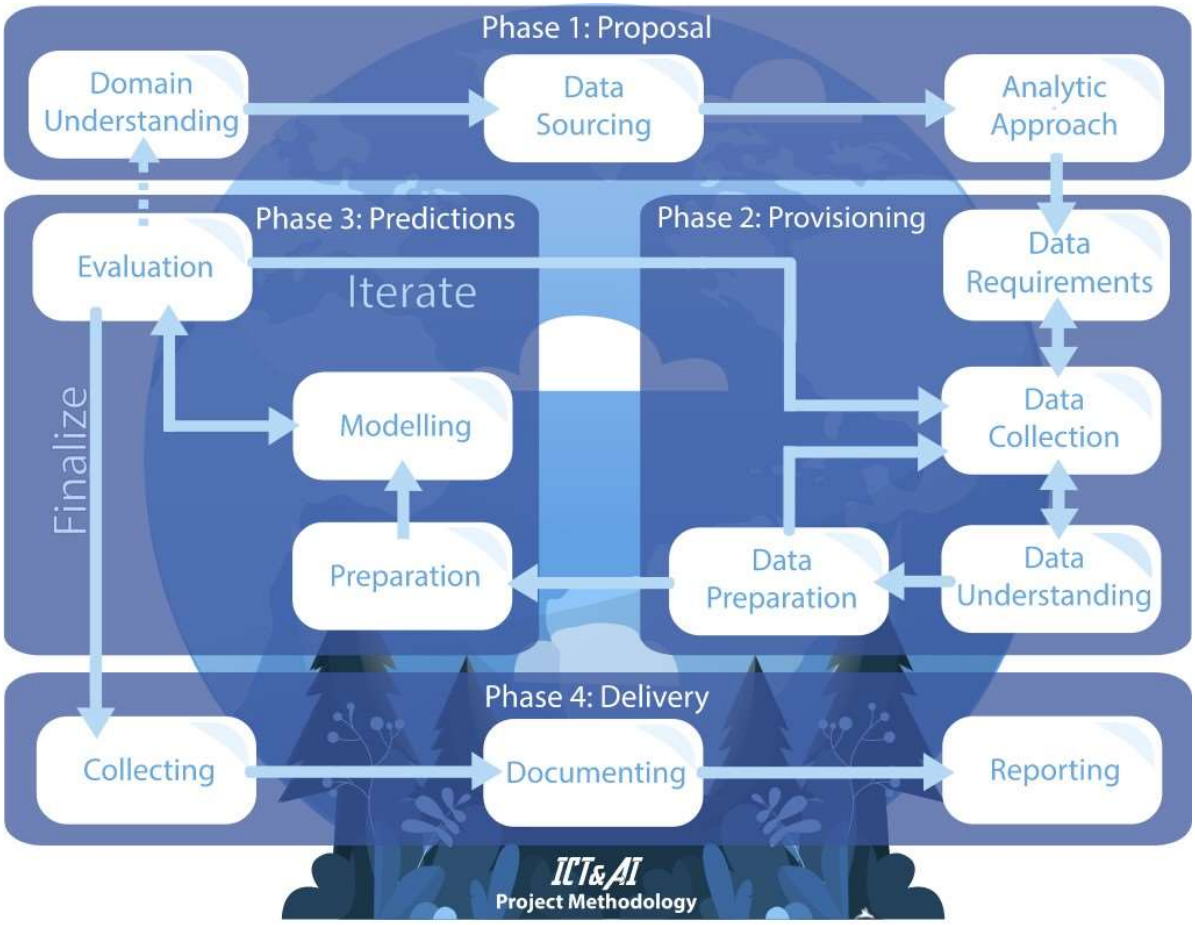
Business Understanding

The current dataset contains about 16500 rows with information about video games with sales greater than 100,000 copies in the world with more focus to North America, Europe, Japan. The dataset tells us number of sales in the continents and world in general, ranking of overall sales, year and platform of the games release ,publisher and genre of the games e.t.c.

In summary, the dataset gives information about sales of video games in continents and in the world at large.

Project Methodology

The project is going to be split into 4. The 4 phases are as follows: Phase 1: Proposal, Phase 2: Provisioning, Phase 3: Predictions, Phase 4: Delivery.



Proposal

In the proposal part I will be working on the initial project proposal and the exploratory data analysis in which the dataset will be explored and also try to find sensible trends in the data that could be used for further process.

Provisioning

In the provisioning phase, there will be a deeper dive into the data trying to expand EDA and prepare the data for modeling. This phase also includes understanding and expanding the data.

Predictions

In the prediction phase I will create a machine learning model and train it with the data that I have at hand to yield a prediction that justifies the project . Different machine learning models will be tested to make sure I can deliver the best possible results.

Delivery

In the delivery phase, I will present the completed project and report to the teachers so as to get feedback.

Project Goal

The main aim is to focus and improve on increasing the sales which could then lead to improving profits.

Stakeholders

- Nick Welman (Semester coach & Internal Stakeholder)
- Konings Hans (Semester coach & Internal Stakeholder)
- Kuijpers Nico(Teacher & Internal Stakeholder)
- Huisman Jose(Teacher & Internal Stakeholder)
- Michael Ebowusim(Developer & Internal Stakeholder)

How are the stakeholders affected?

All Stakeholders listed to produce this technology require its functionality. Effort has been given by each of the stakeholders to make sure the project is known and perceived.

Communication

Communication between the semester coach, teachers and developer will beheld mainly online due to the current situation regarding the pandemic using Microsoft teams as the main platform.

Data Sourcing & Storage

Data Sourcing:

The dataset which I am working with at this was provided by GregorySmith on Kaggle and was generated by a scrape of vgchartz.com.

Data Storage:

For data storage, the data will be stored locally on my PC.

Privacy

This technology does not register personal data, due to the data not having any columns about individual data about people e.g customers, developers of game e.t.c

Data analysis(EDA)

To understand the columns I will be dealing with, below is the data description of these columns.

- Rank - Ranking of overall sales
- Name - The games name
- Platform - Platform of the games release (i.e. PC,PS4, etc.)
- Year - Year of the game's release
- Genre - Genre of the game
- Publisher - Publisher of the game
- NA_Sales - Sales in North America (in millions)
- EU_Sales - Sales in Europe (in millions)
- JP_Sales - Sales in Japan (in millions)
- Other_Sales - Sales in the rest of the world (in millions)
- Global_Sales - Total worldwide sales.

Modelling

On the topic of modelling, I will have discussions with the teacher and then conclude on what algorithms to use. For now, I will just list the one I think:

- Decision Tree
- Support-vector machine
- Linear Regression
- Random Forest

Model Evaluation

I will need to have discussion with my teachers before concluding on this.