

Deliverable 1

1) Team

a) Members

Meghana Chikyala
Veda Samhitha Dyawanapally
Amarnath Reddy Chinthapalli
Poornima Pulakandam
Pavan Kalyan Reddy Madatala

b) Communication plan to include project artifact repository

A communication plan was created to outline how project artifacts were stored and shared among team members. In this case, We have used Google Spaces as the primary platform for discussions, while Google Meet was used for virtual meetings. Additionally, a Git repository will be utilized to share project files and other relevant information. The frequency of group meetings will be based on the project deliverables.

GitHub Repository Link: https://github.com/PavanKalyanReddyM/BDA_Project.git

2) Selection of data:

We have chosen this dataset from kaggle.

DataSet Link:

Choice 1 : **Video Games Sales:**

<https://www.kaggle.com/datasets/gregorut/videogamesales/code?datasetId=284>

Choice 2: **Crime in Context, 1975-2015:**

<https://www.kaggle.com/datasets/marshallproject/crime-rates>

3) Business Problem or Opportunity, Domain Knowledge

a. Video Games Sales:

The dataset we have chosen is from Kaggle where it gives the data related to the Video Game Sales in North America, Europe, Japan, Sales in the rest of the world and Global Sales. We can use this data to derive some useful information like in which Genre of Video Games have high sales, in which place they can release more Video Games and earn more profit and Which Company has highest number of Sales of Video Games and also understand the correlation between different variables that are highly related using Correlation Graph so it gives us which variables in the dataset are closely related so that they can be given more importance.

b. Crime in Context:

The dataset we have chosen is from Kaggle where it gives the data related to the Crime Rates in different states of United States of America and count of Crimes such as Rapes, Violent Crimes, Homicides, Assaults etc. and their reported year. We can use this data to derive some useful information like in which region a high amount of Security is required by creating a visualization on the Cases, which crime is increasing more from 1975 till 2015, in which months the crime rate is more.

4) Research Objectives and Question(s)**a. Video Games Sales:**

- 1) In which genre, The publishers need to focus more to increase the Sales?
- 2) Which genre is the most selling and how competitive is it with other genres?
- 3) Can we predict the genre of a game based on its sales data?
- 4) What genre of Video Games are having high sales?
- 5) Is there a correlation between year of release and global sales?
- 6) Which Video Game is having highest sales so that updated versions can be released for the same to increase Sales?
- 7) Which video game is the most popular in North America, Japan and Europe?
- 8) Which platform is highly preferred to play these Video Games?

b. Crime in Context:

- 1) Which Crime Rate has the highest number of Reported Cases?
- 2) Which State needs to be highly protected based on the number of Reported Cases?
- 3) In which month, the crime rates are higher?
- 4) What is the latest Crime which is occurring based on the reported year?
- 5) What is the correlation between violent crimes and homicides?
- 6) Can we predict the number of robberies that will occur in a certain jurisdiction in a given year based on previous data?
- 7) Is there a correlation between the crime rate per capita and the population size in a given jurisdiction?