VISUALIZING MOVIE DATA

A

PROJECT SYNOPSIS

FOR PROJECT

BACHELOR OF TECHNOLOGY
DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

SUBMITTED BY

DEVYANI GUPTA	**17ETCCS008**
KARTIK KUMAWAT	**17ETCCS020**
RAKSHIT JOSHI	**17ETCCS046**
HIMANSHU TAK	**17ETCCS016**
KHUSHBOO PADDIYAR	**17ETCCS022**

UNDER THE GIIDANCE OF
Mr. Gaurav Kumawat

Techno India NJR Institute of Technology
2020-2021

Visualizing Movie Data

Introduction:

You are a business analyst consultant and your client is a new movie production company looking to make a new movie. The client wants to make sure it's successful to help make a name for the new company. They are relying on you to help understand movie trends to help inform their decision making. They've given you guidance to look into three specific areas:

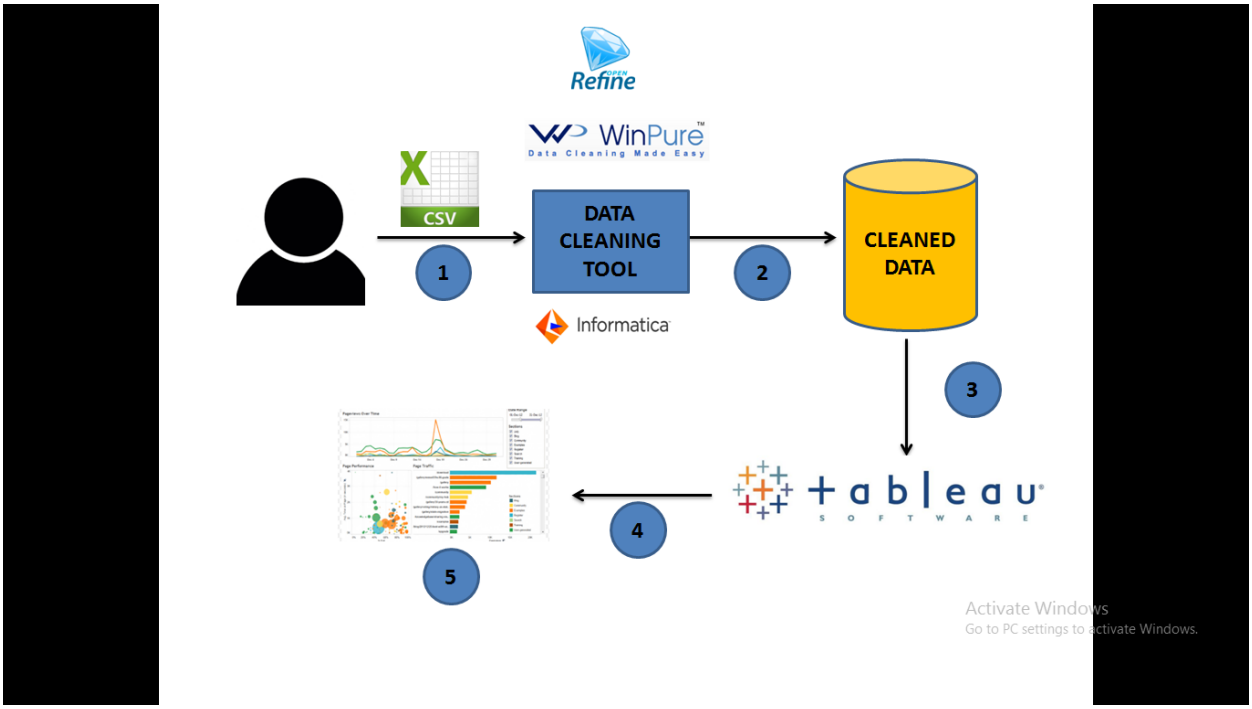
Question 1: How have movie genres changed over time?
Question 2: How do the attributes differ between Universal Pictures and Paramount Pictures?
Question 3: How have movies based on novels performed relative to movies not based on novels?
Question 4: On top of that, the client asked you to explore one other question that you find interesting based on the data provided.

About Data

The Movie Database data can be found in a file called movies.csv at the bottom of this page. To help you understand the data found in movies.csv here is a breakdown of what every field (column) means:
id: Identification number
imdb_id: IMDB identification number
popularity: Relative number of page views on The Movie Database
budget: Budget in USD
revenue: Revenue in USD
original_title: Movie title
cast: list of cast members separated by |, max five actors
homepage: URL for the movie homepage
director: list of directors separated by |, max five directors
tagline: Tagline for the movie
keywords: list of keywords associated with the movie, separated by |, max five keywords
overview: Summary of the plot
runtime: Movie runtime in minutes
genres: list of genres separated by |, max five genres
production_companies: list of production companies separated by |, max five companies

release_date: Original release date
vote_count: Number of votes
vote_average: Average of votes
release_year: Release year
budget_adj: Budget adjusted for inflation, in 2010 US dollars
revenue_adj: Revenue adjusted for inflation, in 2010 US dollars

Architecture Flow:



1. User imports the .CSV file in Data Cleaning Tool like OpenRefine or Winpure or Informatica (Latest Version).

2. The expected output is in .csv file in lossless multidimensional property.

3. The .csv file generated in step-2 will be feeded in Tableau Software

4. Two Visualization per query need to be generated including a Dashboard & Story Visualization.
5. Student can take the help of "Show me" tool bar which suggests the type of graph which can be made over current processed data.

Included Components:

a. Dimensions
b. Measures
c. Marks

Featured Technologies:

a. Hyper – Tableau Data Engine
b. VizQL