

High Level Design (HLD) Entertainer Data Analysis

By: - Avs Jagannath

High Level Design (HLD)

Document Version Control

Date Issued	Version	Description	Author
03 July 2024	1.0	First Version of Complete HLD	Avs Jagannath

Contents

Document Version Control	2
Abstract	4
1 Introduction 1.1 Purpose of this High Design Document 1.2 Scope	5
2 General Description 2.1 Product Perspective & Problem Statement 2.2 Tools Used	6
3.Design Details 3.1 Functional Architecture 3.2 Data Preparation And Integration 3.3 Dashboard Design 3.4 Optimization strategies	7
4 KPIs (Key performance Indicators)	8
5 Deployment	8

Abstract

The Entertainer Data Analysis project aims to provide comprehensive insights into the careers and milestones of influential figures in the entertainment industry over the past century. By analysing data from various sources, this project seeks to uncover trends, patterns, and relationships in the entertainment world, focusing on factors such as breakthrough moments, career longevity, and industry demographics.

The project utilizes a dataset spanning from 1889 to 1988, encompassing 64 entertainers from diverse backgrounds in film, music, and television. Through the use of advanced data processing techniques and interactive visualizations, we aim to shed light on the evolving landscape of the entertainment industry.

Key aspects of the analysis include:

1. Examining the demographics of entertainers and their breakthrough moments across different decades.
2. Investigating the relationship between an entertainer's age, their breakthrough, and subsequent career milestones.
3. Analyzing gender representation and its evolution over time in various entertainment mediums.
4. Exploring trends in career longevity and the factors that contribute to sustained success in the industry.

By leveraging modern business intelligence tools such as Power BI, this project will deliver interactive dashboards that allow users to explore the data dynamically. These visualizations will provide valuable insights for entertainment industry professionals, researchers, and enthusiasts alike, offering a data-driven perspective on the factors that shape successful careers in entertainment.

The findings from this analysis may contribute to a better understanding of industry trends, inform decision-making in talent development, and provide historical context for the current state of the entertainment world.

1. Introduction

1.1 Purpose of this High-Level Design Document

This HLD document aims to provide a comprehensive overview of the Entertainer Data Analysis project, detailing the system architecture, data flow, and key components. It serves as a reference for developers, analysts, and stakeholders involved in the project.

1.2 Scope

The project encompasses data integration from multiple sources, data cleaning and preparation, creation of interactive dashboards, and deployment of the final product using business intelligence tools.

2. General Description

2.1 Product Perspective & Problem Statement

The entertainment industry has a rich history of influential figures who have shaped popular culture. This project aims to analyze data on entertainers from various eras, focusing on their breakthroughs, career milestones, and demographic information. The goal is to uncover trends and insights that reflect the evolution of the entertainment industry over time.

2.2 Tools Used

- Python (Pandas, NumPy) for data preparation



- Power BI for dashboard creation and visualization



- Excel for data storage and initial analysis

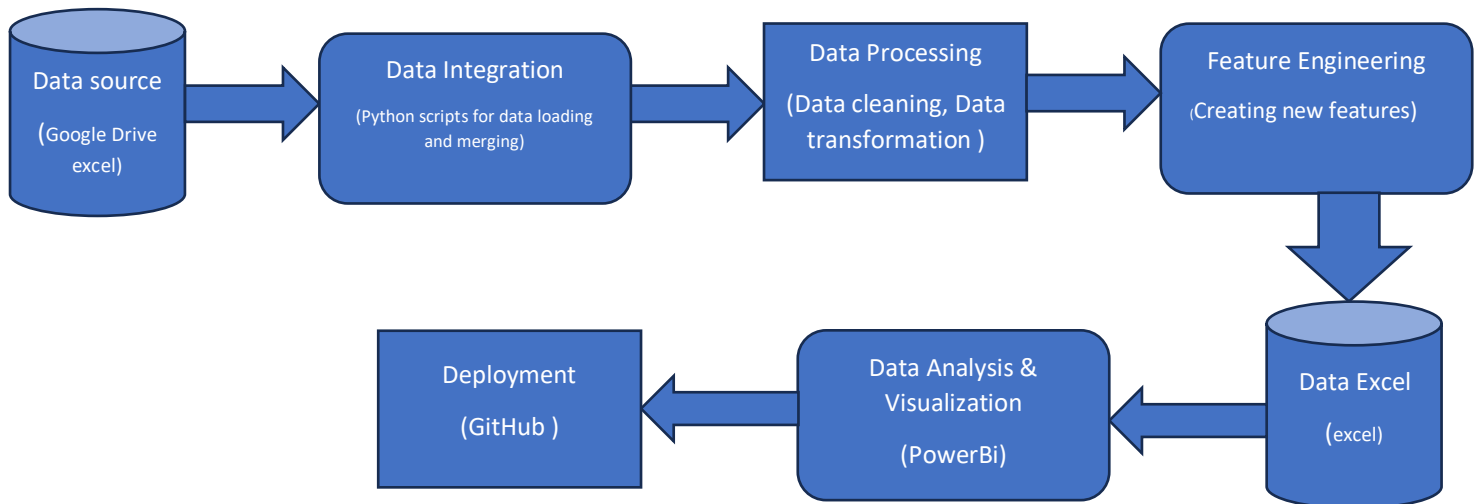


- GitHub for version control and code sharing



3. Design Details

3.1 Functional Architecture



3.2 Data Preparation and Integration

- Data loading from Excel files
- Data cleaning (handling missing values, removing duplicates)
- Data transformation (calculating age, decade of breakthrough)
- Data merging from multiple sources

3.3 Dashboard Design

- Overview of entertainer demographics
- Breakthrough analysis by decade
- Gender representation over time
- Career longevity visualization
- Awards and nominations trends

3.4 Optimization Strategies

- Efficient data loading and processing in Python
- Use of calculated fields in visualization tools
- Implementation of filters and parameters for interactivity
- Proper use of data extracts in Tableau/Power

4. KPIs (Key Performance Indicators)

- Average age at breakthrough by decade
- Gender distribution in breakthroughs over time
- Time between breakthrough and first major award
- Career longevity trends
- Representation of different entertainment mediums (film, music, television)

5. Deployment

- Selection of appropriate deployment option (Power BI Service)
- Data refresh strategy
- User access management
- Performance monitoring and optimization