Entertainer Data Analysis – Project Documentation

**Task #1 – Word Document (Process + Datasets + Dashboard Theme)**

**🔹 Process Followed**

1. **Data Collection**
   * Three datasets provided:
     + **Entertainer – Basic Info**: Contains entertainer’s name, gender, birth year, nationality, profession.
     + **Entertainer – Breakthrough Info**: Contains career breakthrough year, age at breakthrough, breakthrough work.
     + **Entertainer – Last Work Info**: Contains entertainer’s last major work, last work year, awards, and current status (alive/deceased).
2. **Data Preparation**
   * Cleaned missing/null values in Power Query.
   * Standardized date formats (year only).
   * Created **unique Entertainer ID** to merge all datasets.
   * Added calculated columns/measures:
     + **Age at Breakthrough** = Breakthrough Year – Birth Year.
     + **Active Span** = Last Work Year – Breakthrough Year.
     + **Current Age** = TODAY() – Birth Year (if alive).
     + **Is Deceased** flag.
3. **Data Modeling**
   * **Star schema** built in Power BI.
   * Dimensions: *Basic Info, Breakthrough Info, Last Work Info*.
   * Facts: *Career timeline (Breakthrough → Last Work)*.
4. **Dashboard Building**
   * Created multiple report pages:
     + **Overview** (KPIs, breakthrough trends, cumulative insights).
     + **Career Insights** (active spans, median ages, event timelines).
     + **Storyline/Insights** (narrative findings with visuals).
5. **Storyline Creation**
   * Highlighted trends such as:
     + Golden era of breakthroughs (1950s–70s).
     + Female representation growing since 1990s.
     + Entertainers with long active careers (50+ years).

**🔹 Datasets Used**

* **Entertainer – Basic Info.xlsx**
* **Entertainer – Breakthrough Info.xlsx**
* **Entertainer – Last Work Info.xlsx**

**🔹 Dashboard Theme**

* **Theme Name:** “Journey of Entertainers”
* **Color Palette:** Purple gradient with highlight colors (cinema spotlight feel).
* **Visual Focus:** Show an entertainer’s journey from *Birth → Breakthrough → Awards → Last Work → Current Status*.
* **Audience:** Entertainment industry analysts, fans, researchers.

**High-Level Design (HLD)**

* **Objective:** Provide insights into entertainers’ careers from breakthrough to last work.
* **Components:**
  1. **Data Sources:** Excel datasets (Basic, Breakthrough, Last Work).
  2. **ETL Layer:** Power Query for cleaning & transformation.
  3. **Data Model:** Star schema with Entertainer ID as key.
  4. **Analytics Layer:** DAX measures (active span, median age, % female, etc.).
  5. **Visualization Layer:** Power BI dashboards with KPIs, charts, storyline.
  6. **Users:** Students, analysts, entertainment researchers.

**Low-Level Design (LLD)**

* **Tables & Fields:**
  + *Basic Info*: EntertainerID, Name, Gender, BirthYear, Nationality, Profession.
  + *Breakthrough Info*: EntertainerID, BreakthroughYear, BreakthroughWork.
  + *Last Work Info*: EntertainerID, LastWorkYear, LastWork, AwardsCount, IsAlive.
* **Calculated Columns/Measures (DAX):**
  + Age at Breakthrough = BreakthroughYear - BirthYear
  + Active Span = LastWorkYear - BreakthroughYear
  + Current Age = YEAR(TODAY()) - BirthYear
  + Female % = DIVIDE(COUNTROWS(FILTER(BasicInfo, Gender="F")), COUNTROWS(BasicInfo))
* **Relationships:**
  + One-to-many between *Basic Info* and both *Breakthrough Info* & *Last Work Info* via EntertainerID.

**System Architecture**

# 

**Wireframe (Dashboard Layouts)**

**Page 1: Overview**

* Top: KPI Cards → *Total Entertainers, Alive Entertainers, % Female, Median Age, Avg Active Span*.
* Middle Left: Line Chart → *Breakthroughs by Year*.
* Middle Right: Cumulative Chart → *Breakthroughs + Events*.
* Bottom Left: Bar Chart → *Entertainers by Decade*.
* Bottom Right: Table → *Entertainer details*.

**Page 2: Storyline & Insights**

* Top: Text Narrative Box.
* Middle: Entertainer profile cards (Birth, Breakthrough, Awards, Active Span).
* Bottom: Line Chart → *Events by Year (Birth, Breakthrough, Death, Awards, Last Work)*.