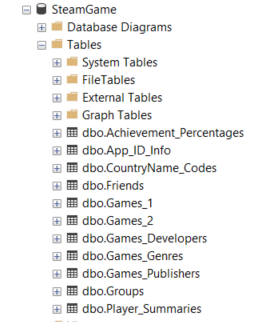
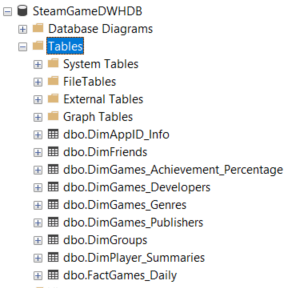
|  |
| --- |
| Savithri B Chaturvedi  **Tableau Data Analysis  on Steam Game Data** |
|  |
| This is a brief summary of how the Tableau Visualization is arrived at from the raw data of Steam Games CSV files |

## Raw Data Transformation

The flat files (CSV) of Steam Game Data (Small Dataset) are first loaded into MSSQL Server as Staging Tables.

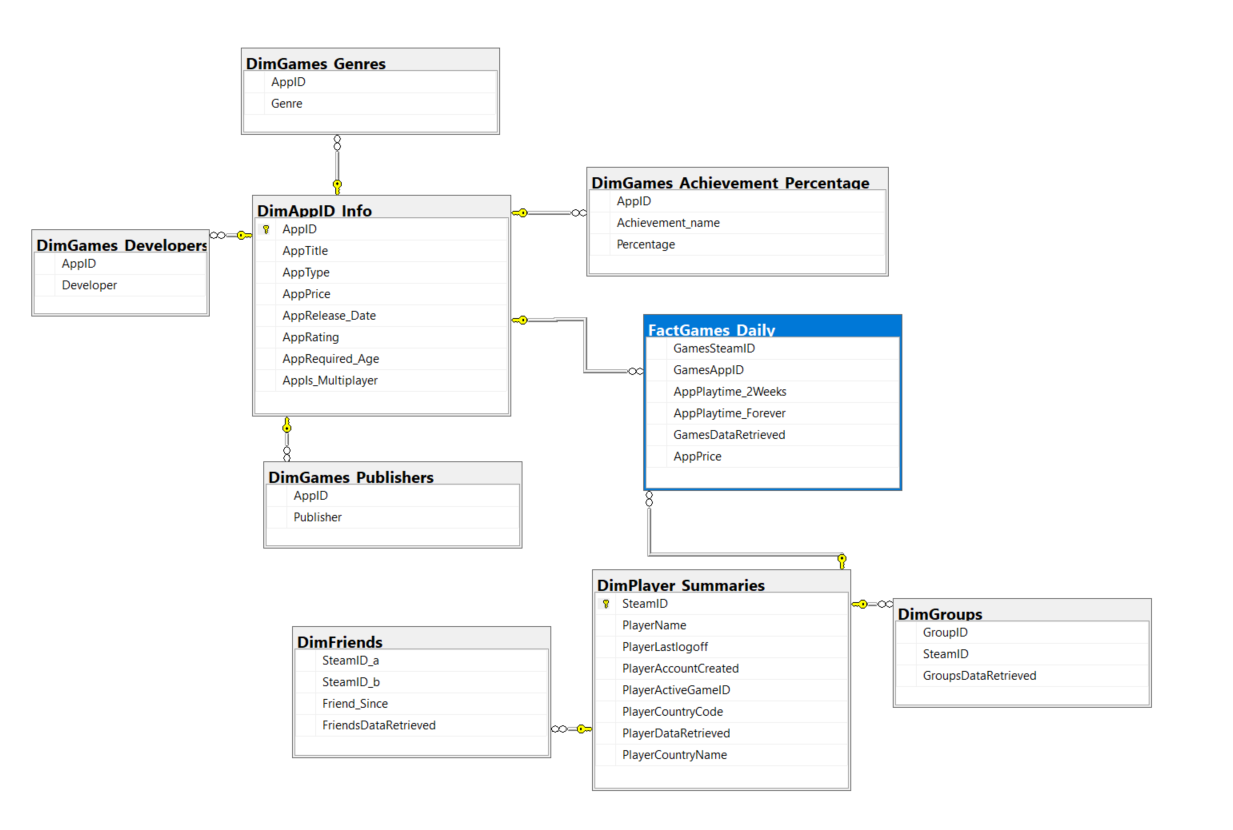
**Staging DB DataWarehouse DB**



After that DataWarehouse DB is created with Dimensions, Facts and identified Primary, ForeignKey constraints. And the data is then cleansed, transformed and loaded as Dimensions & Fact Tables of Data Warehouse DB called - “***SteamGame DWH DB***” using ETL Package Script

(Attached ETL Mapping Document & ETL Package for detailed description of how each field is mapped in Appendix)

**Database Diagram**



**ETL Package to load Datawarehouse Dimensions & Fact Tables from Staging tables**



## Important Note on Data selection

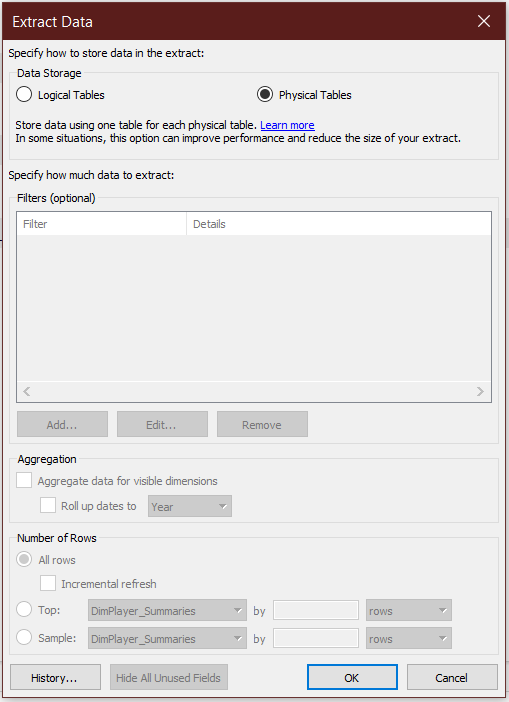
1. Small dataset is used for the entire analysis & visualizations in Tableau as my laptop was freezing with the large dataset
2. Considering the data size and my laptop configuration, I have considered Games\_2 follow up data as FactGames table which is having all the players’ information as that of Game\_1.
3. Since some of the apps had release date, friendship start date etc, among others set to default unix timestamp year1970 and some with 1997 and some with 2002, to get meaningful visualizations only data between ***01-Jan-2003 12-Dec-2012*** has been filtered out and all the visualizations have been created on the same dates.

## Tableau Desktop settings

Connected Tableaue Desktop with MS SQL Datawarehouse DB and then established joins between Dimensions & Fact tables.

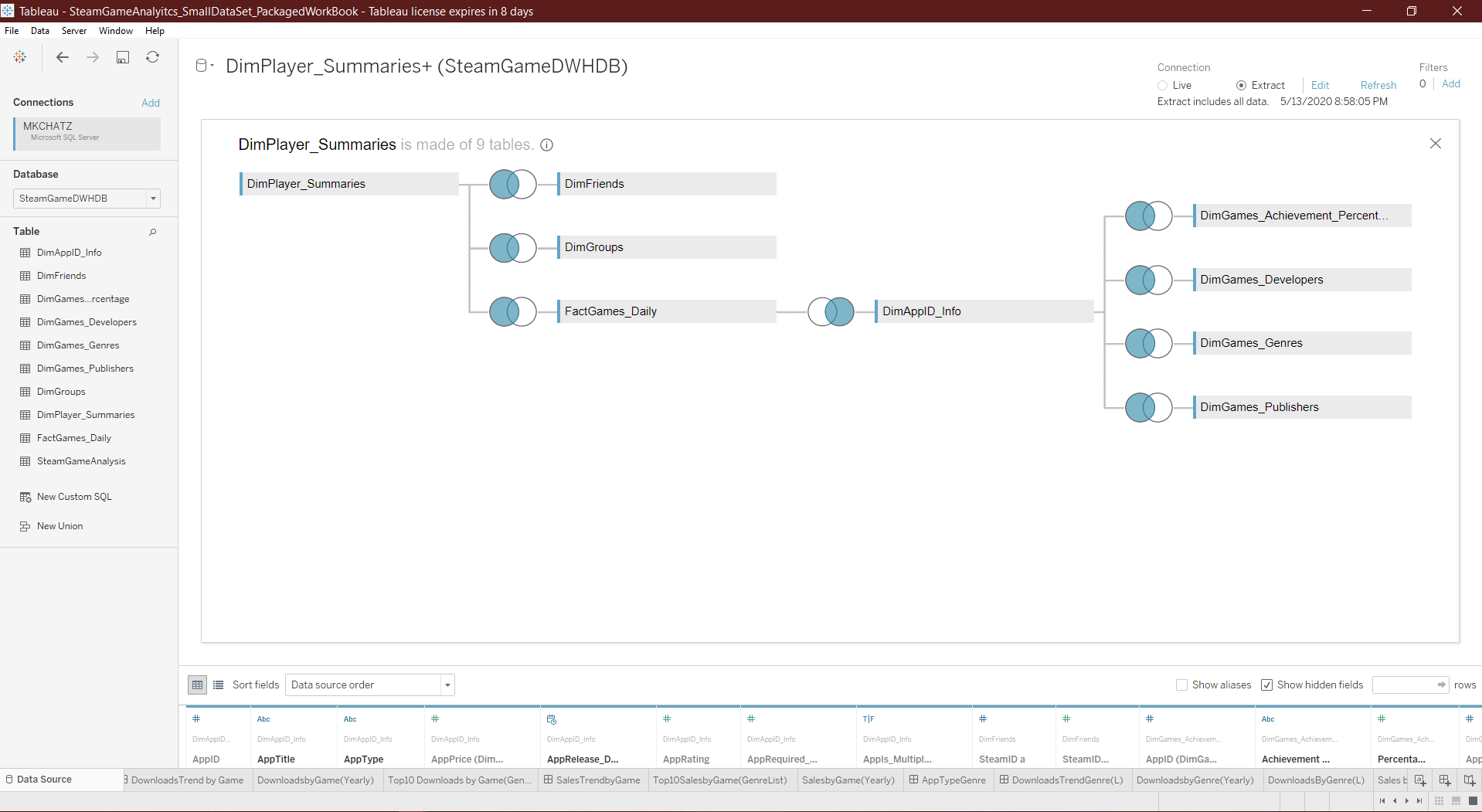
Later I have created Tableau Extract with Physical Table Storage Option.

After the tables are joined, an Extract is made with Physical Tables Storage because the given data is static in nature and does not require Live Refresh option.

**Reasons for choosing Physical Table Storage:**

1. Considerable gain in Performance as the joins are performed on locally available date at query time compared to Live data source.
2. Size of extract is smaller hence it helps in improving
3. Reduced the size of extract file as there are multiple joins involved

**Joins between Dimensions & Fact Tables**



## Tableau Visualizations

**Brief Summary:** The Tableau Work created as part of this data challenge consists of rich data visualizations with multiple story points which give deep insights into how the Games are consumed on Steam and what kind of games are highly popular and liked by the gamers worldwide, how much time is being spent on these games at various levels of filtering, how the game downloads trends are varying with time, further dissected at country, genre, app type, achievements, publishers & developers, avg play time, number of players, top games by downloads and sales,

This dashboard is designed to help various top management executives

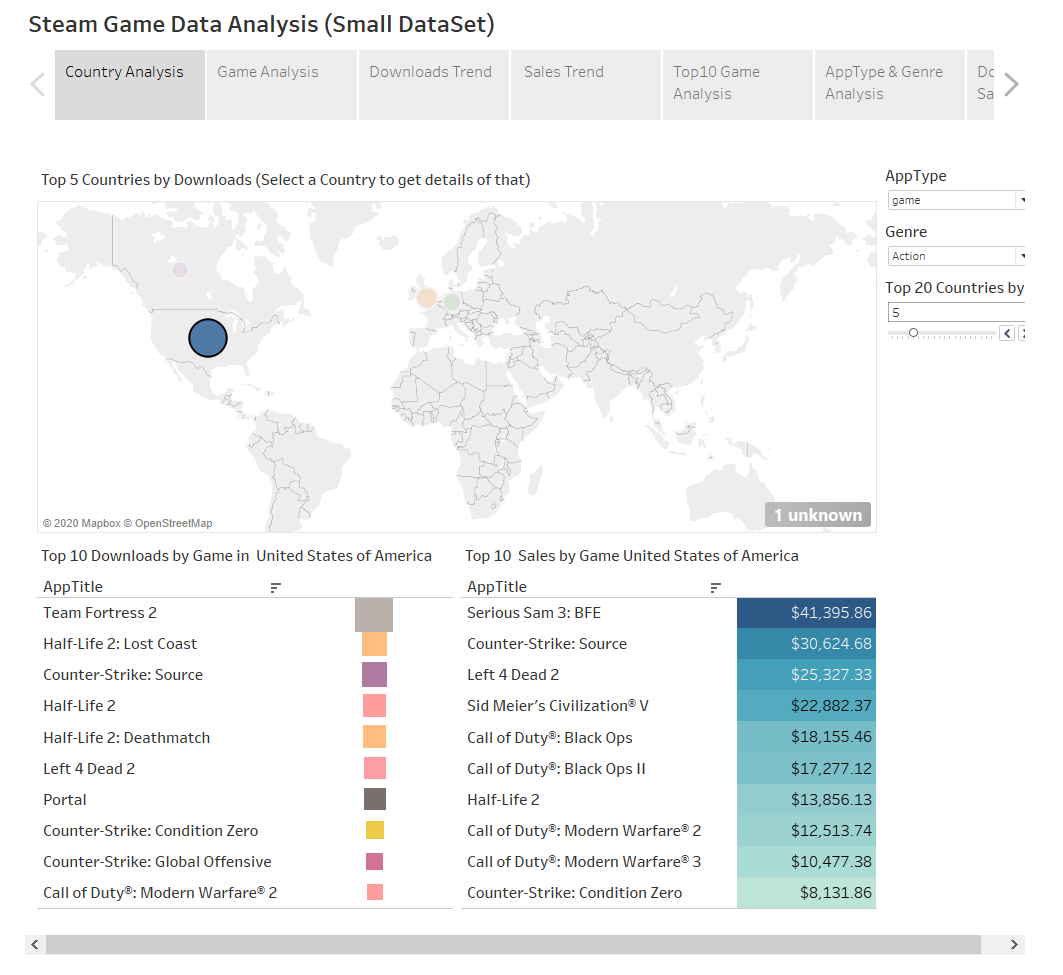
1. **Executive Management Team**: In understanding how Steam is giving revenues and what genres of games are highly profitable
2. **Sales Team:** In evaluating how best to improve their selling abilities by understanding patterns of sales across various geographies and filtered with various genres of top performing, highly engaging, highly downloaded games which brings players back to Steam platform and for more content

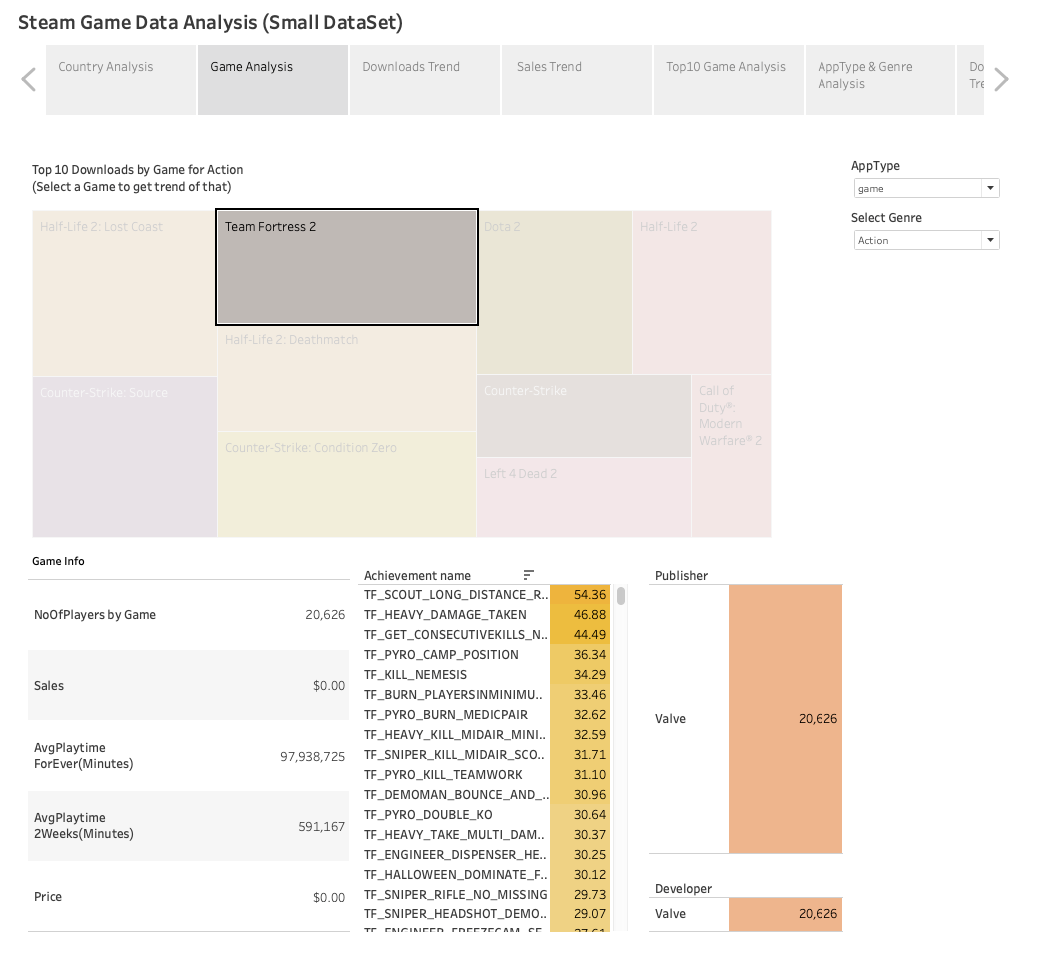
I also wanted to delve deeper into the Friendships, Groups and their impact on how the games are consumed across various geographies and how that is affecting the sales and downloads but due to time and resource constraints I am forced to limit to Smaller Dataset and visualizations related to key aspects such as Sales Revenue, Download patterns, Player engagement with Steam Platform (playtime) etc̣.

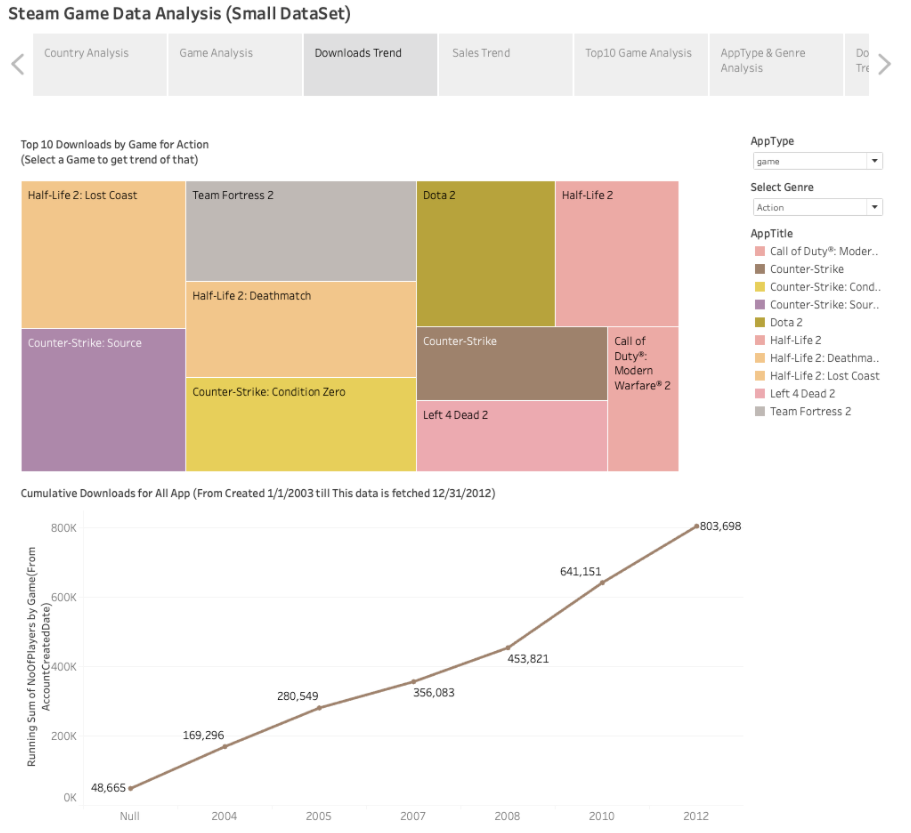
**An overview of Tableau features covered in the Challenge:**

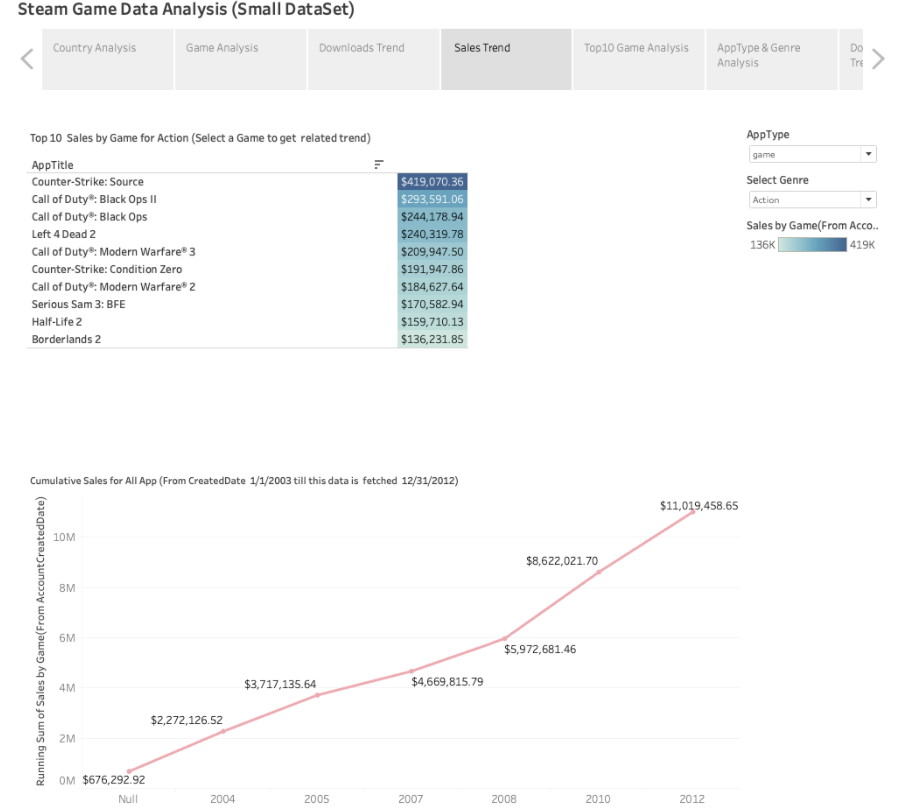
* Used LoDs, Running Total Trend, Dynamic Parameters, Sets, Maps, Dashboard Actions, Context Filters,
* Charts Used: Motion Chart, Bubble Chart, Heat Map, Tree Map, Stacked barchart, Bar Chart, Story Points

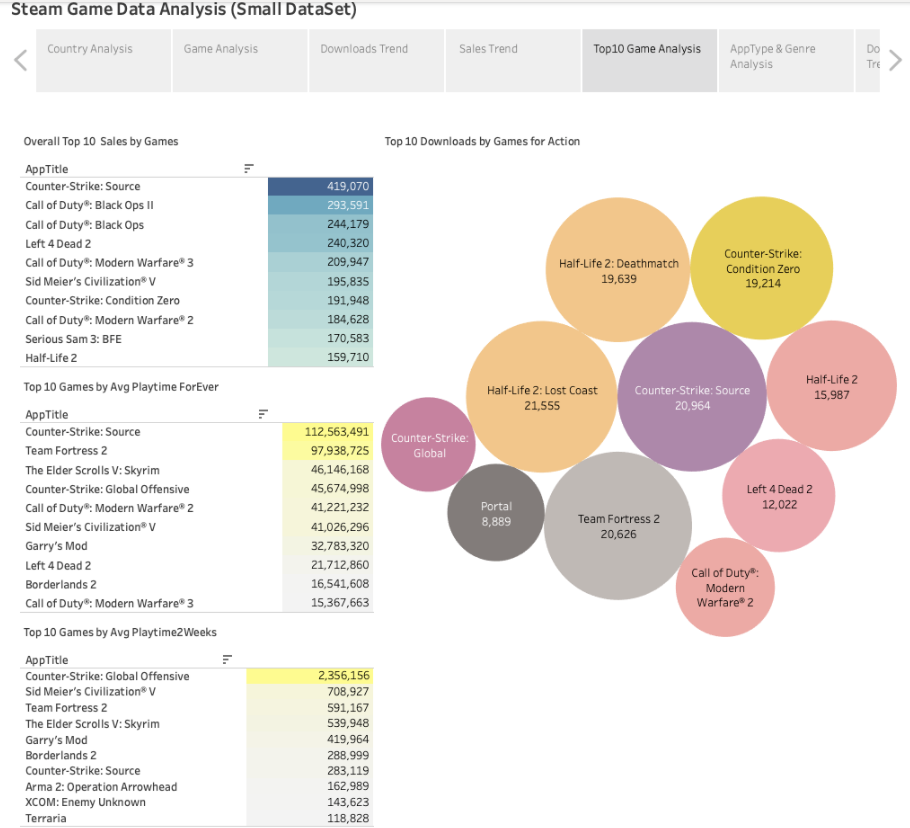
**The Story for the entire workbook is available at the end.**

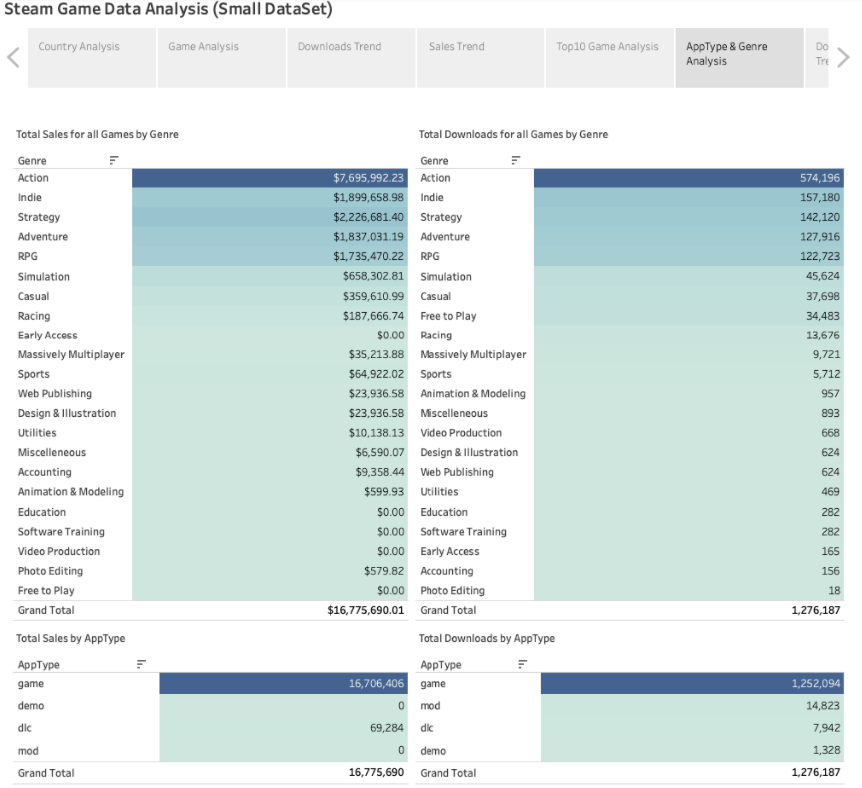


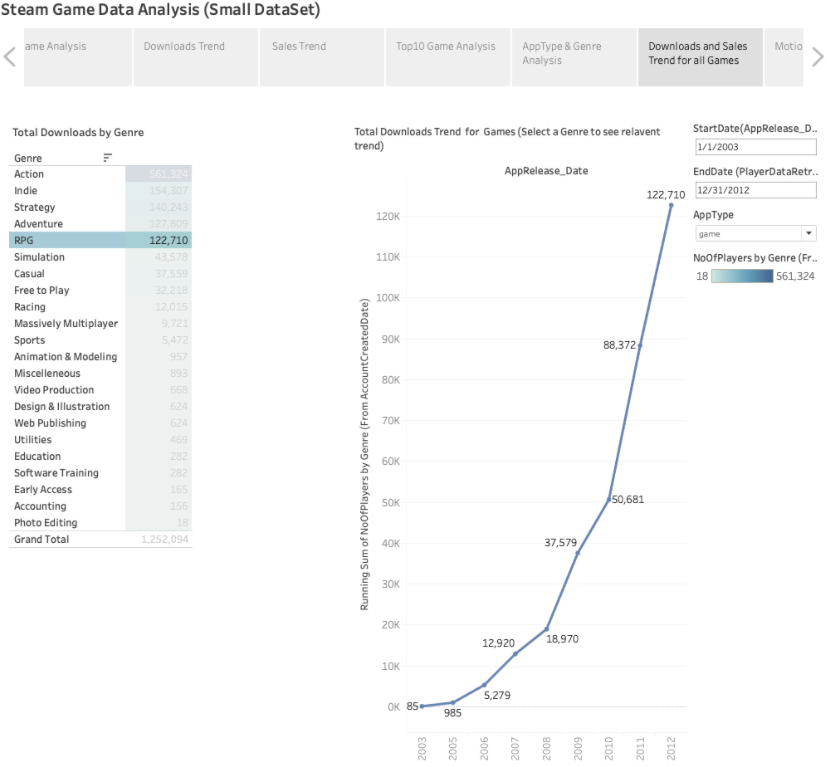












## Appendix

**Data WareHouse DB Creation Script:**



**Data WareHouse DB ETL Mapping Document:**

