

Assignment 2

Irfan Kanat

March 8, 2018

This is an assignment to evaluate your learning in module two. For each learning activity there is a section in this document. Answer each question below the question and create an R Markdown document for submission.

For the rest of the exercises we will use games.csv in the data folder. Import the file into a data.frame called games to begin the exercise.

```
# Import necessary packages
```

```
# Import games.csv from data folder
```

1 - Describe

Visual Inspection

Show the top 6 rows of *games*.

Descriptive Statistics

Use *summary* function to provide basic descriptive statistics for *games*.

Use *str* function to provide an overview of data structure for *games*.

2 - Visualization

Histogram

Plot out the distribution of price variable in a histogram.

Scatter Plot

Plot the relation between avgMin (on average how much time players spent in game) and reviews.

Bar Chart

Plot the frequency of status types in *games*.

3 - Subsetting

Selecting Rows and Columns

Show only the games that have SuperStar status.

Show only the first three rows.

Aggregating

Calculate the average price for each status category, save the results into a data.frame called *gameAvgPrice*.

```
# Using formula interface for aggregate
```

Change the name of the column listPrice in *gameAvgPrice* to avgCategoryPrice.

Combine Datasets

Combine *games* and *gameAvgPrice* so that the average price for category is displayed next to each game. Save the results into *games* data.frame.

I want column avgCategoryPrice to be added to games, in such a way that it shows the average price for the correct status. A game that is a niche player will show 10.87 in avgCategoryPrice for example.

Transformations

Add a new column to *games* named logListPrice, and store log transformed list price in this column. Don't forget log of 0 is meaningless, take the necessary precautions to ensure this does not lead to loss of data.

Missing Data

Show how many rows there are in games.

Exclude any observations that are missing data on reviews column. Save the resulting data.frame in *games*

Show how many rows there are after removal of observations.

Dimension Reduction

WIP