**Applied Statistical Analysis mini project report**

Name – Atharva shevate

Roll no – 2

Div – E(E1)

Srn -202201727

**Introduction**

This project uses a movie rating dataset to formulate relationship between several variables, visualise the patterns of data, data behaviour, analysis of data and create standardized (Z-) scores for several variables, run Frequencies to explore the distributions of several variables, obtain summary statistics for scale variables using Box-plot, visualize the relationship between two scale variables creating scatter plots, three wat & two way cross tabulations to explore the relationship between variables.

**Dataset:**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| srno | movieId | Rating | timestamp | |  |  |
| 1 | 2 | 3.5 | ######## |  |  |  |
| 1 | 29 | 3.5 | ######## |  |  |  |
| 1 | 32 | 3.5 | ######## |  |  |  |
| 1 | 47 | 3.5 | ######## |  |  |  |
| 1 | 50 | 3.5 | ######## |  |  |  |
| 1 | 112 | 3.5 | ######## |  |  |  |
| 1 | 151 | 4 | ######## |  |  |  |
| 1 | 223 | 4 | ######## |  |  |  |
| 1 | 253 | 4 | ######## |  |  |  |
| 1 | 260 | 4 | ######## |  |  |  |
| 1 | 293 | 4 | ######## |  |  |  |
| 1 | 296 | 4 | ######## |  |  |  |
| 1 | 318 | 4 | ######## |  |  |  |
| 1 | 337 | 3.5 |  |  |  |  |
| 1 | 367 | 3.5 | ######## |  |  |  |
| 1 | 541 | 4 | ######## |  |  |  |
| 1 | 589 | 3.5 | ######## |  |  |  |
| 1 | 593 | 3.5 | ######## |  |  |  |
| 1 | 653 | 3 | ######## |  |  |  |
| 1 | 919 | 3.5 | ######## |  |  |  |
| 1 | 924 | 3.5 | ######## |  |  |  |
| 1 | 1009 | 3.5 | ######## |  |  |  |
| 1 | 1036 | 4 | ######## |  |  |  |
| 1 | 1079 | 4 | ######## |  |  |  |
| 1 | 1080 | 3.5 | ######## |  |  |  |
| 1 | 1089 | 3.5 | ######## |  |  |  |
| 1 | 1090 | 4 | ######## |  |  |  |
| 1 | 1097 | 4 | ######## |  |  |  |
| 1 | 1136 | 3.5 | ######## |  |  |  |
| 1 | 1193 | 3.5 | ######## |  |  |  |
| 1 | 1196 | 4.5 | ######## |  |  |  |
| 1 | 1198 | 4.5 | ######## |  |  |  |
| 1 | 1200 | 4 | ######## |  |  |  |
| 1 | 1201 | 3 | ######## |  |  |  |
| 1 | 1208 | 3.5 | ######## |  |  |  |
| 1 | 1214 | 4 | ######## |  |  |  |
| 1 | 1215 | 4 | ######## |  |  |  |
| 1 | 1217 | 3.5 | ######## |  |  |  |
| 1 | 1219 | 4 | ######## |  |  |  |
| 1 | 1222 | 3.5 | ######## |  |  |  |
| 1 | 1240 | 4 | ######## |  |  |  |
| 1 | 1243 | 3 | ######## |  |  |  |
| 1 | 1246 | 3.5 | ######## |  |  |  |
| 1 | 1249 | 4 | ######## |  |  |  |
| 1 | 1258 | 4 | ######## |  |  |  |
| 1 | 1259 | 4 | ######## |  |  |  |
| 1 | 1261 | 3.5 | ######## |  |  |  |
| 1 | 1262 | 3.5 | ######## |  |  |  |
| 1 | 1266 | 4 | ######## |  |  |  |
| 1 | 1278 | 4 | ######## |  |  |  |
| 1 | 1291 | 3.5 | ######## |  |  |  |
| 1 | 1304 | 3 | ######## |  |  |  |
| 1 | 1321 | 4 | ######## |  |  |  |
| 1 | 1333 | 4 | ######## |  |  |  |
| 1 | 1348 | 3.5 | ######## |  |  |  |
| 1 | 1350 | 3.5 | ######## |  |  |  |
| 1 | 1358 | 4 | ######## |  |  |  |
| 1 | 1370 | 3 | ######## |  |  |  |
| 1 | 1374 | 4 | ######## |  |  |  |
| 1 | 1387 | 4 | ######## |  |  |  |
| 1 | 1525 | 3 | ######## |  |  |  |
| 1 | 1584 | 3.5 | ######## |  |  |  |
| 1 | 1750 | 3.5 | ######## |  |  |  |
| 1 | 1848 | 3.5 | ######## |  |  |  |
| 1 | 1920 | 3.5 | ######## |  |  |  |
| 1 | 1967 | 4 | ######## |  |  |  |
| 1 | 1994 | 3.5 | ######## |  |  |  |
| 1 | 1997 | 3.5 | ######## |  |  |  |
| 1 | 2021 | 4 | ######## |  |  |  |
| 1 | 2100 | 4 | ######## |  |  |  |
| 1 | 2118 | 4 | ######## |  |  |  |
| 1 | 2138 | 4 | ######## |  |  |  |
| 1 | 2140 | 4 | ######## |  |  |  |
| 1 | 2143 | 4 | ######## |  |  |  |
| 1 | 2173 | 4 | ######## |  |  |  |
| 1 | 2174 | 4 | ######## |  |  |  |
| 1 | 2193 | 4 | ######## |  |  |  |
| 1 | 2194 | 3.5 | ######## |  |  |  |
| 1 | 2253 | 3.5 | ######## |  |  |  |
| 1 | 2288 | 4 | ######## |  |  |  |
| 1 | 2291 | 4 | ######## |  |  |  |
| 1 | 2542 | 4 | ######## |  |  |  |
| 1 | 2628 | 4 | ######## |  |  |  |
| 1 | 2644 | 3.5 | ######## |  |  |  |
| 1 | 2648 | 3.5 | ######## |  |  |  |
| 1 | 2664 | 3.5 | ######## |  |  |  |
| 1 | 2683 | 3.5 | ######## |  |  |  |
| 1 | 2692 | 3.5 | ######## |  |  |  |
| 1 | 2716 | 3.5 | ######## |  |  |  |
| 1 | 2761 | 3 | ######## |  |  |  |
| 1 | 2762 | 4 | ######## |  |  |  |
| 1 | 2804 | 3.5 | ######## |  |  |  |
| 1 | 2872 | 4 | ######## |  |  |  |
| 1 | 2918 | 3.5 | ######## |  |  |  |
| 1 | 2944 | 4 | ######## |  |  |  |
| 1 | 2947 | 3.5 | ######## |  |  |  |
| 1 | 2959 | 4 | ######## |  |  |  |
| 1 | 2968 | 4 | ######## |  |  |  |
| 1 | 3000 | 3.5 | ######## |  |  |  |
| 1 | 3030 | 3 | ######## |  |  |  |
| 1 | 3037 | 3.5 | ######## |  |  |  |
| 1 | 3081 | 4 | ######## |  |  |  |
| 1 | 3153 | 4 | ######## |  |  |  |
| 1 | 3265 | 3.5 | ######## |  |  |  |
| 1 | 3438 | 3.5 | ######## |  |  |  |
| 1 | 3476 | 3.5 | ######## |  |  |  |
| 1 | 3479 | 4 | ######## |  |  |  |
| 1 | 3489 | 4 | ######## |  |  |  |
| 1 | 3499 | 4 | ######## |  |  |  |
| 1 | 3889 | 4 | ######## |  |  |  |
| 1 | 3932 | 3 | ######## |  |  |  |
| 1 | 3996 | 4 | ######## |  |  |  |
| 1 | 3997 | 3.5 | ######## |  |  |  |
| 1 | 4011 | 4 | ######## |  |  |  |
| 1 | 4027 | 4 | ######## |  |  |  |
| 1 | 4105 | 3.5 | ######## |  |  |  |
| 1 | 4128 | 4 | ######## |  |  |  |
| 1 | 4133 | 3 | ######## |  |  |  |
| 1 | 4226 | 3.5 | ######## |  |  |  |
| 1 | 4306 | 4 | ######## |  |  |  |
| 1 | 4446 | 3.5 | ######## |  |  |  |
| 1 | 4467 | 4 | ######## |  |  |  |
| 1 | 4571 | 4 | ######## |  |  |  |
| 1 | 4720 | 3.5 | ######## |  |  |  |
| 1 | 4754 | 4 | ######## |  |  |  |
| 1 | 4878 | 3.5 | ######## |  |  |  |
| 1 | 4896 | 4 | ######## |  |  |  |
| 1 | 4911 | 4 | ######## |  |  |  |
| 1 | 4915 | 3 | ######## |  |  |  |
| 1 | 4941 | 3.5 | ######## |  |  |  |
| 1 | 4980 | 3.5 | ######## |  |  |  |
| 1 | 4993 | 5 | ######## |  |  |  |
| 1 | 5026 | 4 | ######## |  |  |  |
| 1 | 5039 | 4 | ######## |  |  |  |
| 1 | 5040 | 3 | ######## |  |  |  |
| 1 | 5146 | 3.5 | ######## |  |  |  |
| 1 | 5171 | 4 | ######## |  |  |  |
| 1 | 5540 | 4 | ######## |  |  |  |
| 1 | 5679 | 3.5 | ######## |  |  |  |
| 1 | 5797 | 4 | ######## |  |  |  |
| 1 | 5816 | 4 | ######## |  |  |  |
| 1 | 5898 | 3.5 | ######## |  |  |  |
| 1 | 5952 | 5 | ######## |  |  |  |
| 1 | 5999 | 3.5 | ######## |  |  |  |
| 1 | 6093 | 4 | ######## |  |  |  |
| 1 | 6242 | 3.5 | ######## |  |  |  |
| 1 | 6333 | 4 | ######## |  |  |  |
| 1 | 6502 | 3.5 | ######## |  |  |  |
| 1 | 6539 | 4 | ######## |  |  |  |
| 1 | 6754 | 4 | ######## |  |  |  |
| 1 | 6755 | 3.5 | ######## |  |  |  |
| 1 | 6774 | 4 | ######## |  |  |  |
| 1 | 6807 | 3.5 | ######## |  |  |  |
| 1 | 6834 | 3.5 | ######## |  |  |  |
| 1 | 6888 | 3 | ######## |  |  |  |
| 1 | 7001 | 3.5 | ######## |  |  |  |
| 1 | 7045 | 3.5 | ######## |  |  |  |
| 1 | 7046 | 4 | ######## |  |  |  |
| 1 | 7153 | 5 | ######## |  |  |  |
| 1 | 7164 | 3.5 | ######## |  |  |  |
| 1 | 7247 | 3.5 | ######## |  |  |  |
| 1 | 7387 | 3.5 | ######## |  |  |  |
| 1 | 7389 | 4 | ######## |  |  |  |
| 1 | 7438 | 4 | ######## |  |  |  |
| 1 | 7449 | 3.5 | ######## |  |  |  |
| 1 | 7454 | 4 | ######## |  |  |  |
| 1 | 7482 | 3 | ######## |  |  |  |
| 1 | 7757 | 4 | ######## |  |  |  |
| 1 | 8368 | 4 | ######## |  |  |  |
| 1 | 8482 | 3.5 | ######## |  |  |  |
| 1 | 8507 | 5 | ######## |  |  |  |
| 1 | 8636 | 4.5 | ######## |  |  |  |
| 1 | 8690 | 3.5 | ######## |  |  |  |
| 1 | 8961 | 4 | ######## |  |  |  |
| 1 | 31696 | 4 | ######## |  |  |  |
| 2 | 3 | 4 | ######## |  |  |  |
| 2 | 62 | 5 | ######## |  |  |  |
| 2 | 70 | 5 | ######## |  |  |  |
| 2 | 110 | 4 | ######## |  |  |  |
| 2 | 242 | 3 | ######## |  |  |  |
| 2 | 260 | 5 | ######## |  |  |  |
| 2 | 266 | 5 | ######## |  |  |  |
| 2 | 469 | 3 | ######## |  |  |  |
| 2 | 480 | 5 | ######## |  |  |  |
| 2 | 541 | 5 | ######## |  |  |  |
| 2 | 589 | 5 | ######## |  |  |  |
| 2 | 891 | 2 | ######## |  |  |  |
| 2 | 908 | 4 | ######## |  |  |  |
| 2 | 924 | 5 | ######## |  |  |  |
| 2 | 1121 | 3 | ######## |  |  |  |
| 2 | 1196 | 5 | ######## |  |  |  |
| 2 | 1210 | 5 | ######## |  |  |  |
| 2 | 1214 | 5 | ######## |  |  |  |
| 2 | 1249 | 5 | ######## |  |  |  |
| 2 | 1259 | 5 | ######## |  |  |  |
| 2 | 1270 | 5 | ######## |  |  |  |
| 2 | 1327 | 5 | ######## |  |  |  |
| 2 | 1356 | 5 | ######## |  |  |  |
| 2 | 1544 | 5 | ######## |  |  |  |

**Objective**

The primary objective of this analysis is to explore the dataset, conduct various statistical analyses, and visualize the data to gain insights into the relationships between different variables.

**Progress:**

1. **Data Loading and Examination:** The dataset was successfully loaded into the R environment, and its structure was examined to ensure data integrity and quality.
2. **Descriptive Statistics:** Summary statistics, including mean, median, minimum, maximum, and standard deviation, were calculated for each of the numerical variables, namely SepalLengthCm, SepalWidthCm, PetalLengthCm, and PetalWidthCm.
3. **Visual Analysis:** Several types of visualizations were created to understand the distributions and relationships within the dataset. Histograms, box plots, and scatter plots were used to depict the distributions of each variable and the relationships between different pairs of variables.
4. **Inferential Statistics:** A chi-square test was performed to analyze the relationship between the categorical variable and other variables in the dataset.
5. **Z-Scores Calculation:** Z-scores were computed for the numerical variables to standardize the data and identify patterns or outliers.
6. **Frequencies and Distributions:** Frequency tables and distribution plots were generated to explore the occurrence and distribution of different variable values.

**Next Steps and Conclusion**

The initial analysis and visualization of the Iris dataset have provided valuable insights into the

characteristics of the dataset. Further analysis could involve advanced statistical modeling,

such as clustering or classification algorithms, to classify new observations based on the patterns

identified in the dataset.

In the next phase, we will focus on advanced data modeling techniques to develop predictive models and gain a deeper understanding of the underlying patterns and relationships in the dataset. Further research will involve implementing machine learning algorithms to classify the different variables