

Data Analysis Report

Introduction

This section introduces the dataset and the purpose of this analysis.

Data Overview

Load the dataset and display basic statistics and missing values.

First few rows of the data set

	Unnamed: 0	Id	Name	Original_name	Popularity \
0	0	94954	Hazbin Hotel	Hazbin Hotel	277.966
1	1	1396	Breaking Bad	Breaking Bad	492.705
2	2	126308	Sh�� gun	Sh�� gun	1403.663
3	3	94605	Arcane	Arcane	89.432
4	4	37854	One Piece	��f-��f��f"��f��f, ��f	164.547

	First_air_date	Vote_average	Vote_count
0	18-01-2024	9.037	738
1	20-01-2008	8.906	13285
2	27-02-2024	8.807	279
3	06-11-2021	8.749	3668
4	20-10-1999	8.700	4368

Shape of the data set

(1925, 8)

Description of data set

Get summary statistics of numerical variables

```
print(df.describe())
```

```
Unnamed: 0      int64
Id              int64
Name            object
Original_name    object
Popularity       float64
First_air_date   object
Vote_average     float64
Vote_count       int64
dtype: object
```

Amount of missing values in the dataset

```

Unnamed: 0      0
Id              0
Name            0
Original_name   0
Popularity      0
First_air_date  0
Vote_average    0
Vote_count      0
dtype: int64

```

For further analysis purposes all none existing Numerical data has been filled with a mean value

Dataset without NaN valeus

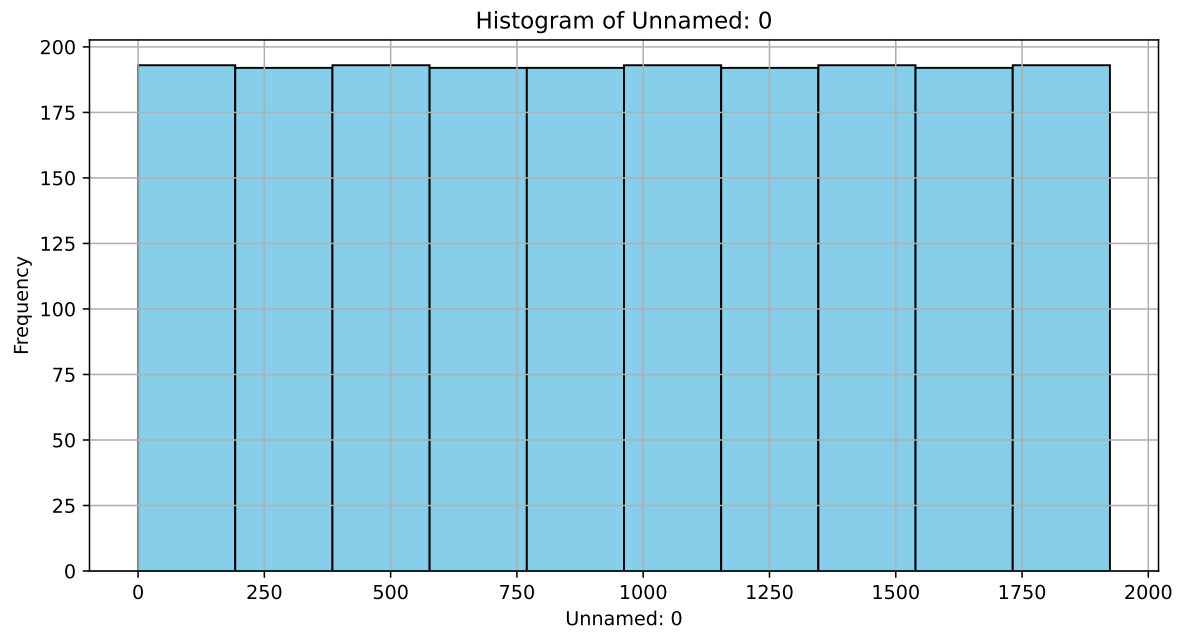
Here is how the first few rows of the cleaned dataset look:

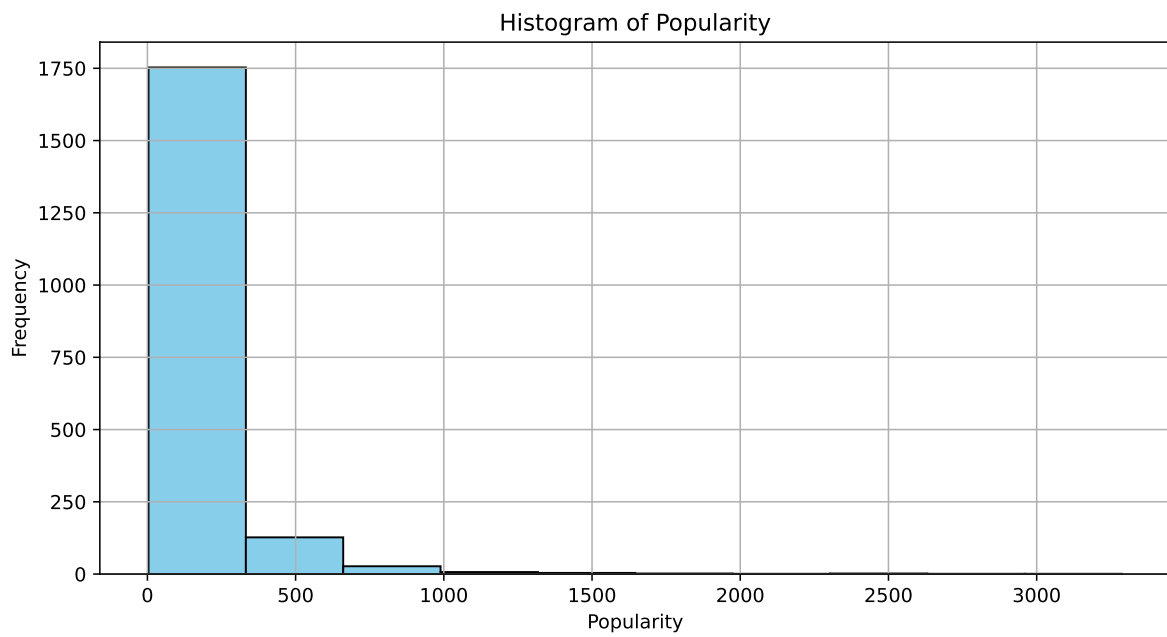
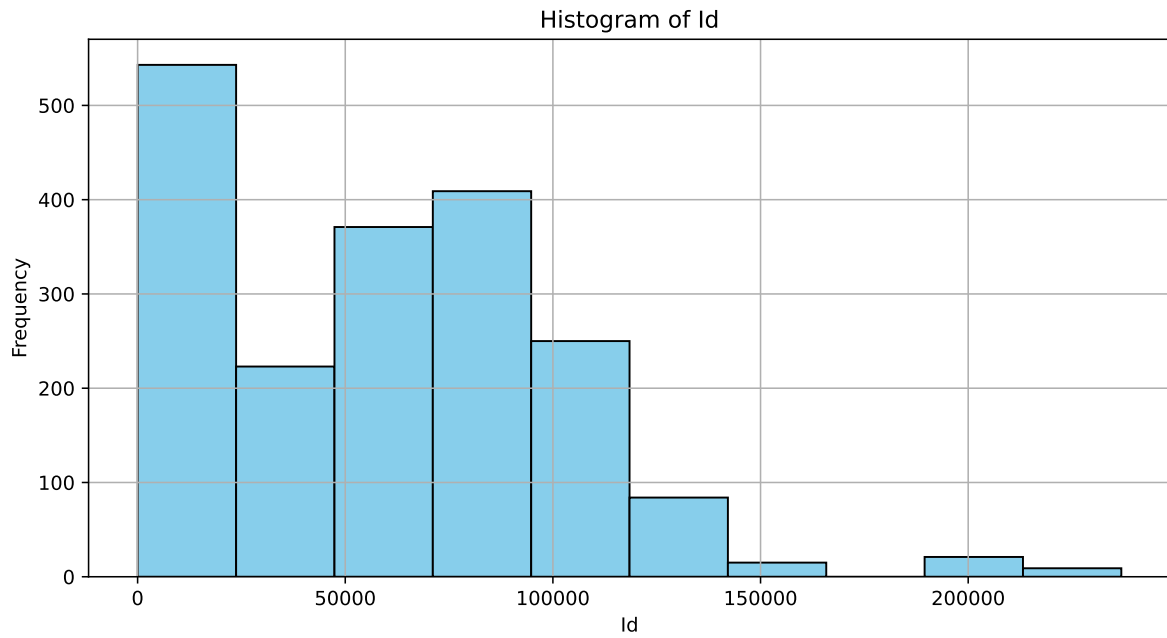
[illegible]

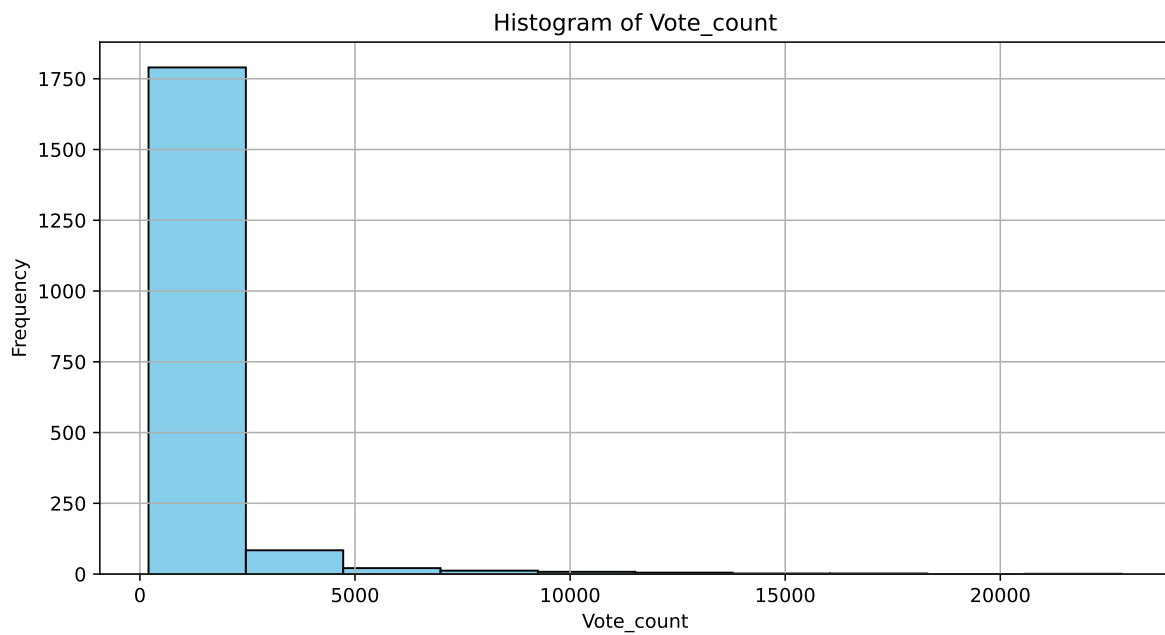
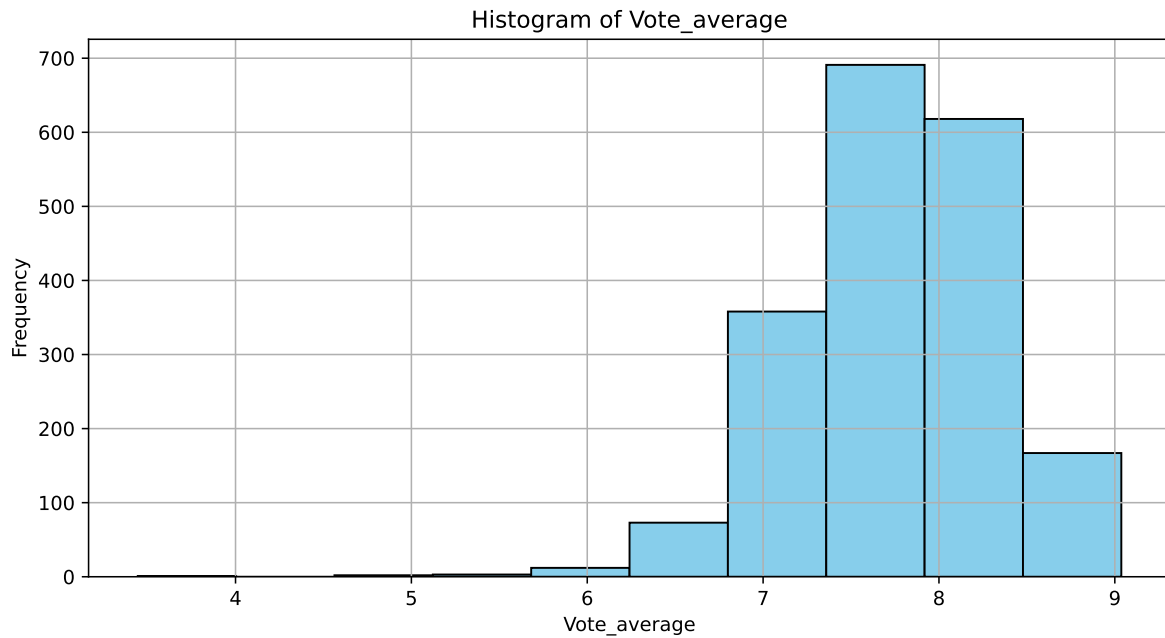
[1925 rows x 8 columns]

In[36]:

Histograms over the data







Correlation matrix over features

Correlation matrix:

Unnamed: 0	Id	Popularity	Vote_average	Vote_count
------------	----	------------	--------------	------------

Unnamed: 0	1.000000	-0.048941	-0.078448	-0.962084	-0.266851
Id	-0.048941	1.000000	-0.224112	0.044500	-0.070166
Popularity	-0.078448	-0.224112	1.000000	0.061174	0.297244
Vote_average	-0.962084	0.044500	0.061174	1.000000	0.252177
Vote_count	-0.266851	-0.070166	0.297244	0.252177	1.000000

