



File Edit View Run Kernel Settings Help

Code

```
def categorize_vote_average(df, col):
    labels = ['non_popular', 'below_average', 'average', 'popular']

    # Ensure the column is numeric
    df[col] = pd.to_numeric(df[col], errors='coerce')

    # Get statistical edges
    stats = df[col].describe() # Computes min, 25%, 50%, 75%, max
    edges = [stats['min'], stats['25%'], stats['50%'], stats['75%']

    # Replace the existing column with categorical values
    df[col] = pd.cut(df[col], bins=edges, labels=labels, duplicate

    return df

# Example usage
df = categorize_vote_average(df, 'Vote_Average')
print(df.head()) # Display the updated dataframe
```

	Release_Date	Title	Popularity	Vote_Count
0	2021.0	Spider-Man: No Way Home	5083.954	8940
1	2022.0	The Batman	3827.658	1151
2	2022.0	No Exit	2618.087	122
3	2021.0	Encanto	2402.201	5076
4	2021.0	The King's Man	1895.511	1793

	Vote_Average	Genre
0	popular	Action, Adventure, Science Fiction
1	popular	Crime, Mystery, Thriller
2	below_average	Thriller
3	popular	Animation, Comedy, Family, Fantasy
4	average	Action, Adventure, Thriller, War

[36]: df.head()