Drugs, Side Effects and Medical Condition Analysis & Machine Learning

Exploratory Data Analysis | Regression & Classification | ML Model

Objective

- Perform exploratory data analysis (EDA) on a drug side effects dataset

- Understand how features like medical condition, drug class, and alcohol interaction influence overall side effect severity

- Build regression models to predict numeric overall_score values

- Build classification models to categorize side effect severity into Low, Medium, and High

- Evaluate and compare model performance using metrics like accuracy, F1-score, and R^2

- Visualize key patterns and relationships using charts and confusion matrices

Dataset Overview

Loaded using pandas

```python
# Reading the dataset
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import xgboost as xgb
import seaborn as sns
from sklearn.ensemble import GradientBoostingRegressor
from sklearn.metrics import mean_squared_error, r2_score
from sklearn.model_selection import GridSearchCV
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import MinMaxScaler, StandardScaler
from sklearn.svm import SVR
from scipy.stats import ttest_ind
from sklearn.linear_model import SGDRegressor
from sklearn.tree import DecisionTreeRegressor
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error
from keras.models import Sequential
from keras.layers import Dense, LSTM
from statsmodels.tsa.seasonal import seasonal_decompose
from statsmodels.tsa.stattools import adfuller, kpss
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
from statsmodels.tsa.arima.model import ARIMA
from wordcloud import wordcloud
from sklearn.pipeline import make_pipeline
import warnings
import statsmodels.api as sm
from statsmodels.api import OLS, add_constant
warnings.filterwarnings('ignore')
%matplotlib inline
sns.set_style("darkgrid")
df=pd.read_csv(r"C:\Users\hp\OneDrive - subho\Desktop\UNIFIED\Drugs, Side Effects and Medical Condit
print(df)
```

Duplicates removed

```
False    489
Name: count, dtype: int64
Shape After deleting duplicate values: (489, 14)
```

Missing values checked

```
drug_name                       0
medical_condition               0
side_effects                    0
generic_name                    0
drug_classes                    0
brand_names                     0
activity                        0
rx_otc                          0
pregnancy_category              0
alcohol                         0
related_drugs                   0
medical_condition_description   0
rating                          0
no_of_reviews                   0
dtype: int64
```

Number of rows and columns present

	drug_name	medical_condition	side_effects	generic_name	drug_classes	brand_names	activity	rx_otc	pregnancy_category	csa	alcohol	
0	doxycycline	Acne	(hives, difficult breathing, swelling in your ...	doxycycline	Miscellaneous antimalarials, Tetracyclines	Acticlate, Adoxa CK, Adoxa Pak, Adoxa TT, Alod...	87%	Rx	D	N	X	https://www.r
1	spironolactone	Acne	hives ; difficulty breathing; swelling of your...	spironolactone	Aldosterone receptor antagonists, Potassium-sp...	Aldactone, CaroSpir	82%	Rx	C	N	X	https://www.dru
2	minocycline	Acne	skin rash, fever, swollen glands, flu-like sym...	minocycline	Tetracyclines	Dynacin, Minocin, Minolira, Solodyn, Ximino, V...	48%	Rx	D	N	NaN	https://www.r
3	Accutane	Acne	problems with your vision or hearing; muscle o...	isotretinoin (oral)	Miscellaneous antineoplastics, Miscellaneous u...	NaN	41%	Rx	X	N	X	https://www.dr
4	clindamycin	Acne	hives ; difficult breathing; swelling of your ...	clindamycin topical	Topical acne agents, Vaginal anti-infectives	Cleocin T, Clindacin ETZ, Clindacin P, Clindag...	39%	Rx	B	N	NaN	https://www.dr

Unique values for each column

```
#Finging unique values for each column
#To understand which column is categorical and which one is Continuous
df.nunique()
```

```
drug_name                        488
medical_condition                 45
side_effects                     485
generic_name                     404
drug_classes                     171
brand_names                      485
activity                          65
rx_otc                             3
pregnancy_category                 7
alcohol                            2
related_drugs                    285
medical_condition_description     45
rating                            77
no_of_reviews                    167
dtype: int64
```

Medical Condition with the Most Unique Drugs

- Observation: Acne leads with the highest number of unique drugs (39, 7.98%), followed by Hypertension (34, 6.95%) and Pain (30, 6.13%), indicating a greater variety of drug options available for treating these conditions.

medical_condition	Unique Drug Count	Percentage (%)
Acne	39	7.98
Hypertension	34	6.95
Pain	30	6.13
Diabetes (Type 2)	26	5.32
Diabetes (Type 1)	21	4.29
Hayfever	20	4.09
AIDS/HIV	19	3.89
Migraine	19	3.89
Insomnia	18	3.68
Osteoarthritis	18	3.68
Asthma	16	3.27
Colds & Flu	16	3.27
Constipation	16	3.27
ADHD	16	3.27
Bipolar Disorder	15	3.07
Anxiety	15	3.07
Psoriasis	13	2.66
Osteoporosis	12	2.45
Depression	12	2.45
GERD (Heartburn)	11	2.25
Eczema	9	1.84
Incontinence	8	1.64
UTI	7	1.43
Cholesterol	7	1.43
Bronchitis	7	1.43
Schizophrenia	6	1.23
Rheumatoid Arthritis	6	1.23
Angina	6	1.23
Weight Loss	5	1.02
Diarrhea	5	1.02
Erectile Dysfunction	4	0.82
Hypothyroidism	4	0.82
Covid 19	4	0.82
Pneumonia	4	0.82
Seizures	4	0.82
Gout	3	0.61
COPD	3	0.61
Cancer	2	0.41
Menopause	2	0.41
Swine Flu	2	0.41

Average Overall Score by Medical Condition

- Observation: Erectile Dysfunction has the highest average overall score (17.90), followed by Anxiety (15.29) and Migraine (12.45), indicating higher user satisfaction or effectiveness for drugs treating these conditions. Conversely, Cancer (2.54) and Stroke (3.01) have the lowest average scores, suggesting less favorable outcomes or user experiences.

medical_condition	
Erectile Dysfunction	17.903223
Anxiety	15.286861
Migraine	12.453164
Bipolar Disorder	12.015586
ADHD	11.501066
Hair Loss	11.390861
Allergies	10.952945
Depression	10.789750
Acne	10.552724
Gout	10.455992
Weight Loss	10.139824
Rheumatoid Arthritis	9.942147
Pain	9.855803
Hypothyroidism	9.069934
Alzheimer's	9.022941
Insomnia	8.924618
Osteoarthritis	8.619112
IBD (Bowel)	8.585466
Diabetes (Type 2)	8.453100
Menopause	8.188088
Constipation	8.056834
Schizophrenia	8.035870
Eczema	7.796846
Bronchitis	7.618559
GERD (Heartburn)	7.233710
Asthma	7.100740
Cholesterol	7.038890
Hypertension	6.916565
Incontinence	6.906186
Psoriasis	6.832721
Covid 19	6.570835
Hayfever	6.312902
Swine Flu	6.172601
Seizures	6.076605
AIDS/HIV	6.018004
UTI	5.778710
Osteoporosis	5.067987
Diabetes (Type 1)	4.970566
COPD	4.479751
Diarrhea	4.255404

Average User Rating by Medical Condition

- Observation: Stroke, Eczema, and Swine Flu have the highest average overall drug scores, suggesting high perceived effectiveness. In contrast, Cholesterol, Allergies, and Hair Loss have lower average scores, indicating potential gaps in treatment satisfaction.

| medical_condition	
Stroke	10.000000
Eczema	8.811111
Swine Flu	8.650000
Gout	8.366667
Erectile Dysfunction	8.350000
Schizophrenia	8.283333
Alzheimer's	8.100000
Anxiety	8.026667
Weight Loss	7.880000
Migraine	7.852632
Asthma	7.806250
GERD (Heartburn)	7.754545
Pneumonia	7.750000
Depression	7.658333
Diabetes (Type 1)	7.595238
Seizures	7.400000
Menopause	7.400000
Acne	7.384615
ADHD	7.350000
IBD (Bowel)	7.300000
Bipolar Disorder	7.180000
Psoriasis	7.153846
Constipation	7.150000
Pain	7.030000
Rheumatoid Arthritis	6.850000
Angina	6.833333
Hypothyroidism	6.650000
Insomnia	6.505556
Osteoarthritis	6.327778
UTI	6.171429
Bronchitis	6.114286
Hayfever	6.090000
AIDS/HIV	6.052632
Diarrhea	6.000000
Colds & Flu	5.931250
Diabetes (Type 2)	5.857692
Hypertension	5.779412
Incontinence	5.725000
Osteoporosis	5.633333
Cancer	5.500000
Covid 19	5.350000

Drugs with the Highest and Lowest Number of Reviews

- Observation : Bisacodyl has the highest number of reviews (1,357), indicating widespread usage or user engagement. In contrast, Neomycin has only 1 review, suggesting limited feedback or niche usage.

```
Drug with the highest number of reviews: bisacodyl , no of reviews: 1357.0
Drug with the lowest number of reviews: neomycin , no of reviews: 1.0
```

Most Commonly Reported Side Effects

- Observation : The most frequently reported side effects include top lip and tongue irritation, stomach pain, serious side effects, and swelling of the face, lips, and difficulty breathing. These side effects highlight key areas for patient monitoring and caution when prescribing or using these medications

Most Common side effects listed across all drugs
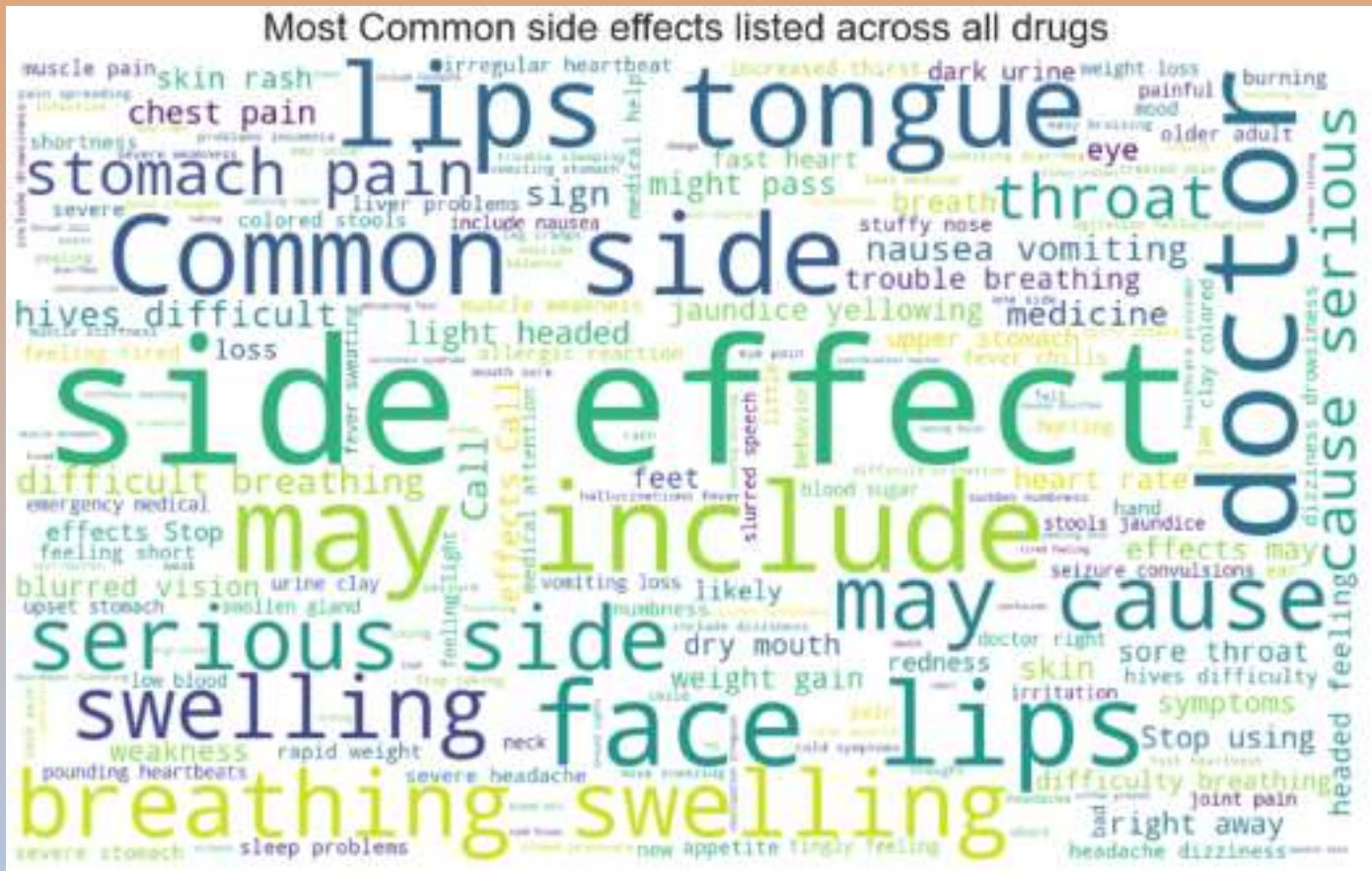
Correlation Between Number of Reviews and User Rating

- Observation: There is no significant correlation between the number of reviews and drug rating ($R^2 = 0.002$, $p = 0.326$). This suggests that review volume does not meaningfully influence or reflect user ratings.

Regression Equation: rating = -0.0 × no of reviews + 6.9
R-squared: 0.002
P-value: 0.326

Correlation between Number of Reviews and Rating

Rating
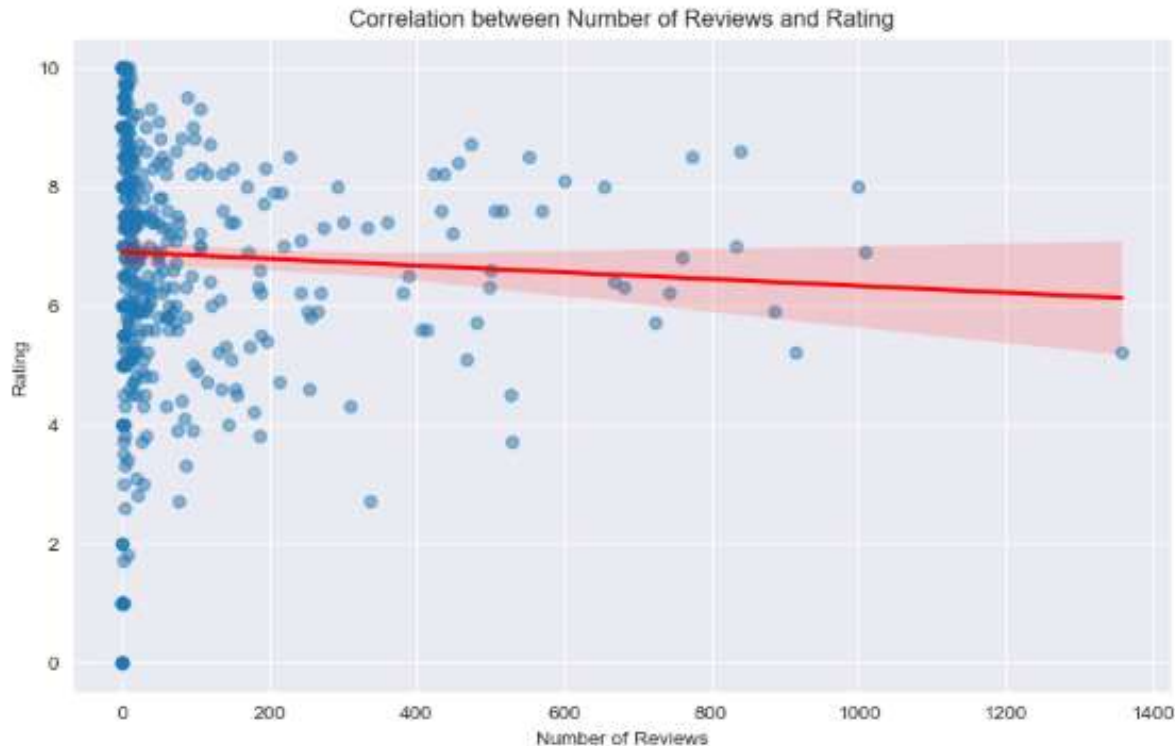
Number of Reviews

The correlation is weak — number of reviews does not explain much of the variation in rating.
The predictor 'number of reviews' is not statistically significant — the relationship with rating is likely due to random variation.

Distribution of Average Ratings by Prescription Type

- Observation: Drugs available as both Rx/OTC have the highest average rating (6.97), slightly outperforming Rx-only (6.84) and OTC-only (6.69) drugs. This indicates that dual-availability drugs may offer a balance of accessibility and effectiveness, leading to higher user satisfaction.

```
rx_otc
Rx/OTC      6.972857
Rx          6.842037
OTC         6.688889
Name: rating, dtype: float64
```

Top 5 Most Common Generic Drugs

- Observation : The top generic drugs include Isotretinoin (oral), Benzoyl Peroxide and Clindamycin (topical), Pramlintide, Metformin and Sitagliptin, and Diphenhydramine. These drugs represent common treatments for conditions ranging from acne and diabetes to allergies and weight management.


Top 5 Most Common Generic Drugs
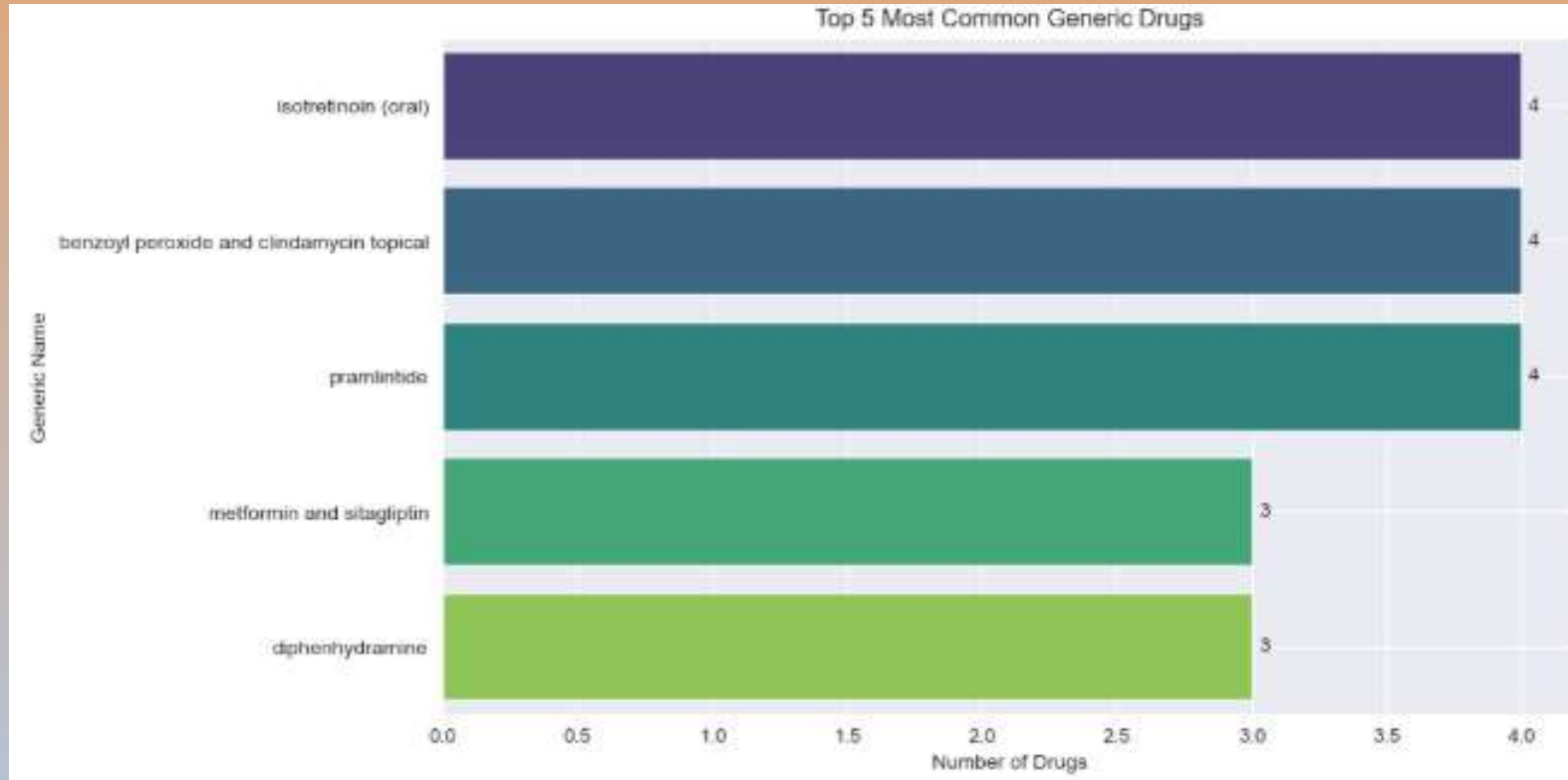
Top 5 Most Common Brand-Name Drugs

- Observation : Among the top 5 most frequently occurring brand-name drugs, Bronkaid, Sudafed PE Severe Cold, and two other brands are each associated with 2 drugs, while the fifth brand group has 1 drug linked to it.

Top 5 Most Common Brands
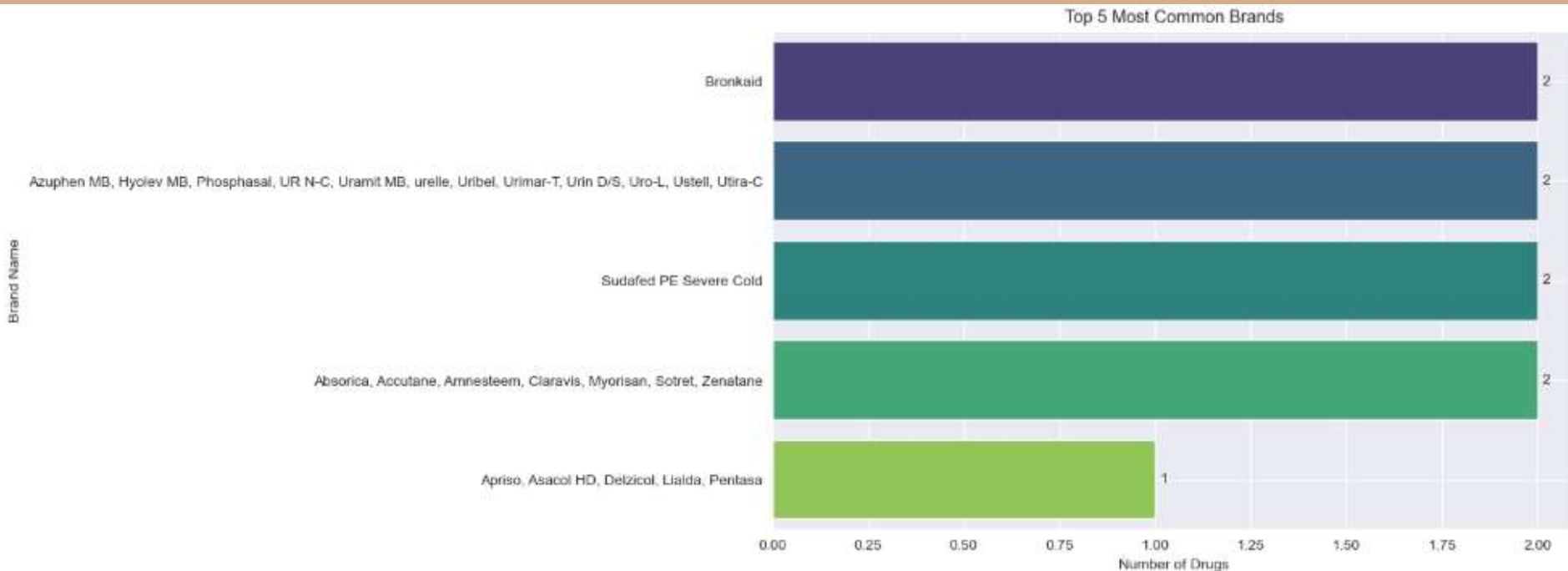
Top 10 Most common drug classes

• Observation: The most frequently occurring drug classes include Upper Respiratory Combinations, Insulin, Laxatives, Antidiabetic Combinations, and Topical Acne Agents. These reflect common treatment areas such as respiratory issues, diabetes management, gastrointestinal health, and dermatological conditions.

Top 10 Most Common drug classes
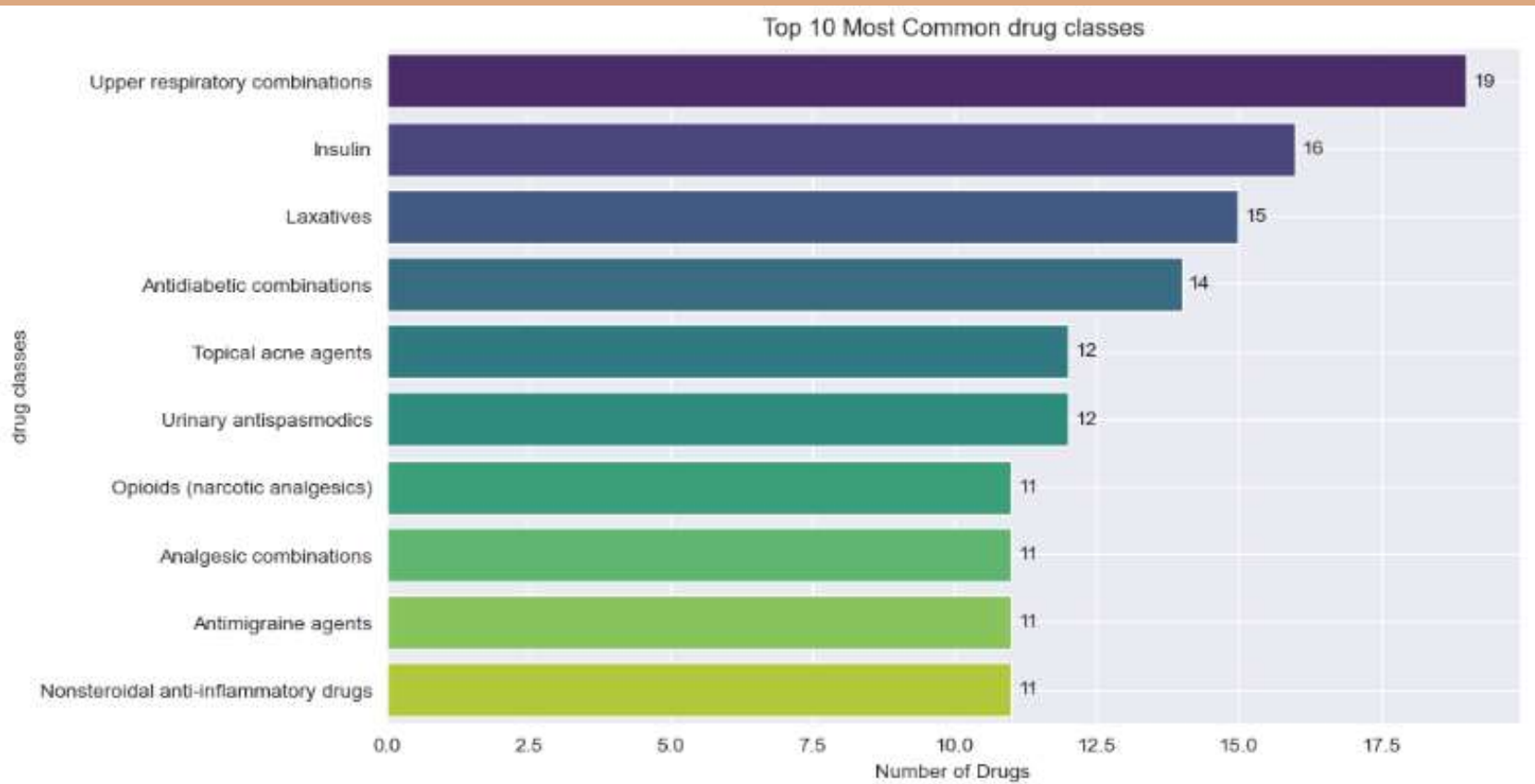
Drug Count by Medical Condition (Low-Risk Pregnancy)

• Observation: Diabetes (Type 1 and Type 2) tops the list with the highest number of low-risk pregnancy drugs (12 each), followed by AIDS/HIV and Hayfever. This indicates a strong pharmaceutical focus on managing chronic and infectious conditions safely during pregnancy.

	pregnancy_category	medical_condition	Number of Drugs
11	B	Diabetes (Type 1)	12
12	B	Diabetes (Type 2)	12
2	B	AIDS/HIV	8
15	B	Hayfever	6
14	B	GERD (Heartburn)	5
13	B	Erectile Dysfunction	4
0	A	Hypothyroidism	4
21	B	Pain	3
3	B	Acne	3
8	B	Bronchitis	3
6	B	Asthma	2
23	B	Psoriasis	2
22	B	Pneumonia	2
16	B	Hypertension	2
18	B	Incontinence	2
19	B	Insomnia	2
7	B	Bipolar Disorder	1
26	B	Seizures	1
25	B	Schizophrenia	1
24	B	Rheumatoid Arthritis	1
17	B	IBD (Bowel)	1
20	B	Osteoarthritis	1
1	B	ADHD	1
4	B	Allergies	1
5	B	Angina	1
10	B	Constipation	1
9	B	Colds & Flu	1
27	B	UTI	1

Drug Count by Medical Condition (High-Risk Pregnancy)

- Observation : Hypertension and Pain lead with the highest number of high-risk pregnancy drugs (22 and 21 respectively), followed by Acne, Hayfever, and ADHD. This reflects the challenges in safely managing both chronic and common conditions during pregnancy with potential risks.

|	pregnancy_category	medical_condition	Number of Drugs
45	D	Hypertension	22
29	C	Pain	21
2	C	Acne	16
22	C	Hayfever	14
0	C	ADHD	12
16	C	Diabetes (Type 2)	11
49	X	Acne	11
23	C	Hypertension	10
6	C	Asthma	10
26	C	Migraine	10
12	C	Constipation	9
7	C	Bipolar Disorder	9
15	C	Diabetes (Type 1)	9
18	C	Eczema	8
11	C	Colds & Flu	8
25	C	Insomnia	8
27	C	Osteoarthritis	8
41	D	Anxiety	7
14	C	Depression	7
31	C	Psoriasis	6
24	C	Incontinence	6
5	C	Anxiety	6
36	C	UTI	5
55	X	Osteoporosis	5
1	C	AIDS/HIV	5
28	C	Osteoporosis	5
39	D	Acne	5
58	X	Weight Loss	4
8	C	Bronchitis	4
19	C	GERD (Heartburn)	4
4	C	Angina	3
10	C	Cholesterol	3
42	D	Bipolar Disorder	3
32	C	Rheumatoid Arthritis	3
43	D	Cancer	2
48	D	Seizures	2
17	C	Diarrhea	2
53	X	Migraine	2

Drugs with Alcohol Restriction

- Observation : The majority of drugs with alcohol restrictions are for Diabetes (Type 2), Pain, Diabetes (Type 1), and Hypertension, together making up over 34% of restricted drugs. This highlights the importance of caution in alcohol consumption for patients managing chronic conditions.

medical_condition	Drug Count	Percentage (%)
Diabetes (Type 2)	26	9.96
Pain	24	9.20
Diabetes (Type 1)	20	7.66
Hypertension	20	7.66
Insomnia	18	6.90
Acne	17	6.51
Bipolar Disorder	15	5.75
Anxiety	15	5.75
Colds & Flu	14	5.36
ADHD	13	4.98
Depression	11	4.21
Migraine	10	3.83
Osteoarthritis	9	3.45
Hayfever	7	2.68
Incontinence	6	2.30
Schizophrenia	5	1.92
UTI	5	1.92
Erectile Dysfunction	4	1.53
Cholesterol	3	1.15
Seizures	3	1.15
Weight Loss	3	1.15
Angina	3	1.15
Covid 19	2	0.77
Allergies	1	0.38
COPD	1	0.38
Cancer	1	0.38
AIDS/HIV	1	0.38
Psoriasis	1	0.38
Bronchitis	1	0.38
Stroke	1	0.38
GERD (Heartburn)	1	0.38

Relationship Between Drug Activity % and User Rating

- Observation : There is no significant correlation between drug activity percentage and user rating (R^2 = 0.001, p = 0.614). This indicates that drug activity levels do not meaningfully affect user satisfaction or rating.

Regression Equation: rating = -0.0 × activity + 6.88
R-squared: 0.001
P-value: 0.614

Correlation between activity and rating
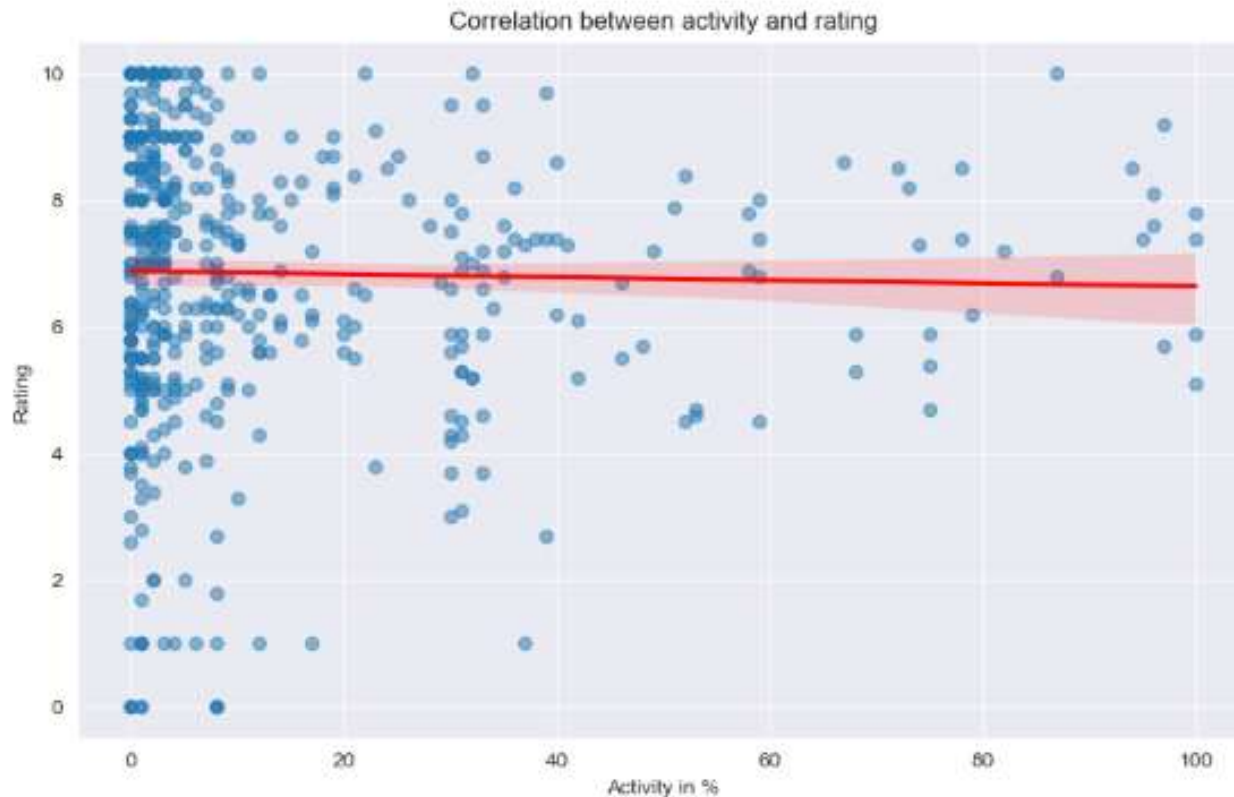
The correlation is weak — activity does not explain much of the variation in rating.
Activity is not statistically significant — the relationship with rating is likely due to random variation.

Do Prescription Drugs (Rx) Have Higher Overall Scores?

- Observation :Statistical testing shows Rx drugs have a significantly higher mean overall score (8.89) compared to non-Rx drugs (6.38), with a strong t-statistic of 4.877 and p-value of 0. We reject the null hypothesis, confirming that prescription drugs tend to be rated higher than non-prescription drugs..

```
T-statistic: 4.877
P-value: 0.0
Mean overall score for Rx drugs: 8.89
Mean overall score for non-Rx drugs: 6.38
Reject the null hypothesis: Rx categorized drugs have significantly different overall score.
Rx drugs tend to have a higher overall score.
```

Do Higher Pregnancy Risk Drugs Have Higher Overall Scores?

- Observation : Hypothesis testing indicates that higher pregnancy risk drugs (C, D, X) have a significantly higher mean overall score (8.87) compared to lower risk drugs (7.44), with a t-statistic of 3.017 and p-value of 0.0027. We reject the null hypothesis, confirming that drugs with higher pregnancy risk tend to be rated more favorably.

```
T-statistic: 3.017
P-value: 0.0027
Mean overall score for higher pregnancy risk drugs: 8.87
Mean overall score for lower pregnancy risk drugs: 7.44
Reject the null hypothesis: higher pregnancy risk drugs have significantly different overall score.
higher pregnancy risk drugs tend to have a higher overall score.
```

Machine Learning – Feature engineering (regression model)

```python
# Machine Learning

# Predicting the overall_score based on the features in the dataset

from sklearn.model_selection import train_test_split
from sklearn.preprocessing import LabelEncoder

# Label Encoding categorical features
label_encoder = LabelEncoder()

df['medical_condition'] = label_encoder.fit_transform(df['medical_condition'])
df['drug_classes'] = label_encoder.fit_transform(df['drug_classes'])
df['rx_otc'] = label_encoder.fit_transform(df['rx_otc'])
df['pregnancy_category'] = label_encoder.fit_transform(df['pregnancy_category'])
df['alcohol'] = label_encoder.fit_transform(df['alcohol'])


# Define features and target variable
X = df[['medical_condition', 'drug_classes', 'rx_otc', 'pregnancy_category', 'alcohol', 'activity']]
y = df['overall_score']

# Handle missing values
X.fillna(X.median(), inplace=True)
y.fillna(y.median(), inplace=True)

# Split the data
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)
```

Models

REGRESSION RANDOM FOREST MODEL

```
RMSE (Random forest): 4.43010066398132
MSE (Random forest): 19.625791893007737
R² (Random forest): 0.314275100304532
```

REGRESSION SVR MODEL

```
RMSE (SVR): 5.788622085678783
MSE (SVR): 33.50814565080819
R² (SVR): -0.17077414968248705
```

REGRESSION SGD MODEL

```
RMSE (SGD): 4.76774111543045
MSE (SGD): 22.731355343765994
R² (SGD): 0.20576675590861304
```

REGRESSION MODEL DECISSION TREE

```
RMSE (Decission tree): 6.304175719435847
MSE (Decission tree): 39.74263150152448
R² (Decission tree): -0.3886069998391415
```

## REGRESSION GRADIENT BOOSTING REGRESSOR MODEL

```
RMSE (GBR): 4.442036070120998
MSE (GBR): 19.731684448256
R² (GBR): 0.3105752158758184
```

REGRESSION xGBoost MODEL

```
RMSE (XGBoost): 4.407102019535134
MSE (XGBoost): 19.422548210590655
R² (XGBoost): 0.3213764317819561
```

Feature importance in XGBoost

- Observation : The most important features influencing the XGBoost model are Activity, Drug Classes, Medical Condition, Pregnancy Category, Rx/OTC status, and Alcohol Restriction. These features play a key role in predicting drug-related outcomes and user ratings effectively.

Feature Importance in XGBoost
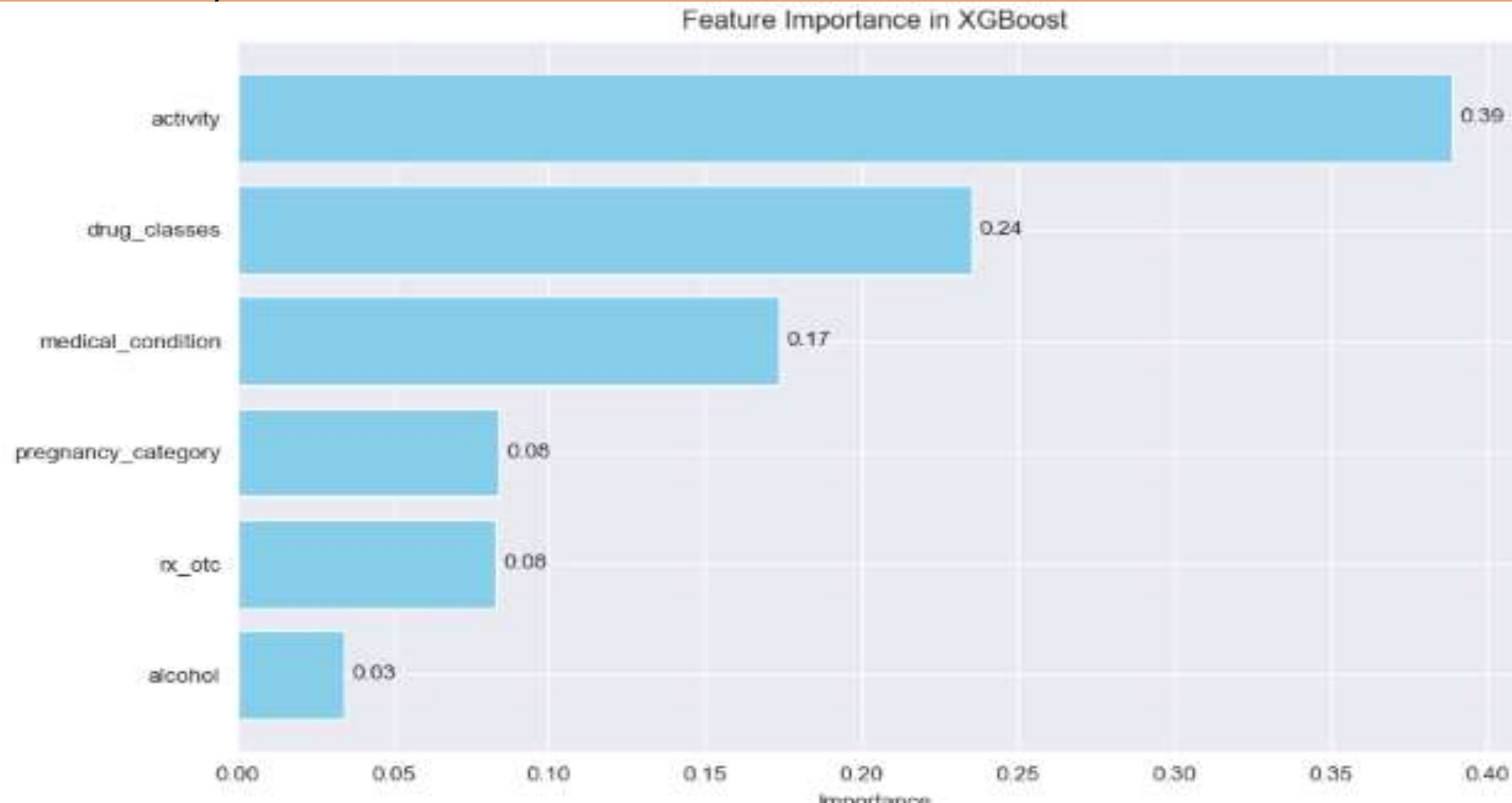
Models

CLASSIFICATION XGBoost MODEL

```
Accuracy: 40.14%
Balanced Accuracy: 0.4013605442176871
F1 Score (Macro): 0.40214285714285714
Classification Report:
              precision    recall  f1-score   support

           0       0.57      0.57      0.57        49
           1       0.38      0.37      0.38        49
           2       0.25      0.27      0.26        49

    accuracy                           0.40       147
   macro avg       0.40      0.40      0.40       147
weighted avg       0.40      0.40      0.40       147

Confusion Matrix:
[[28  8 13]
 [ 6 18 25]
 [15 21 13]]
```

CLASSIFICATION SVC MODEL

```
Accuracy: 42.18%
Balanced Accuracy: 0.4217687074829932
F1 Score (Macro): 0.41633850086985436
Classification Report:
              precision    recall  f1-score   support

           0       0.50      0.49      0.49        49
           1       0.42      0.51      0.46        49
           2       0.33      0.27      0.30        49

    accuracy                           0.42       147
   macro avg       0.42      0.42      0.42       147
weighted avg       0.42      0.42      0.42       147

Confusion Matrix:
[[24 11 14]
 [12 25 12]
 [12 24 13]]
```

CLASSIFICATION SGD MODEL

```
Accuracy: 34.01%
Balanced Accuracy: 0.3401360544217687
F1 Score (Macro): 0.30240506136796713
Classification Report:
              precision    recall  f1-score   support

           0       0.36      0.47      0.41        49
           1       0.23      0.06      0.10        49
           2       0.34      0.49      0.40        49

    accuracy                           0.34       147
   macro avg       0.31      0.34      0.30       147
weighted avg       0.31      0.34      0.30       147

Confusion Matrix:
[[23  4 22]
 [22  3 24]
 [19  6 24]]
```

CLASSIFICATION GRADIENT BOOSTING MODEL

```
Accuracy: 38.78%
Balanced Accuracy: 0.38775510204081637
F1 Score (Macro): 0.39043533185715096
Classification Report:
              precision    recall  f1-score   support

           0       0.50      0.47      0.48        49
           1       0.40      0.39      0.40        49
           2       0.28      0.31      0.29        49

    accuracy                           0.39       147
   macro avg       0.39      0.39      0.39       147
weighted avg       0.39      0.39      0.39       147

Confusion Matrix:
[[23  9 17]
 [ 8 19 22]
 [15 19 15]]
```

CLASSIFICATION RANDOM FOREST MODEL

```
Class distribution: {0: 163, 1: 163, 2: 163}
Accuracy: 42.18%
Balanced Accuracy: 0.4217687074829932
F1 Score (Macro): 0.4235544217687075
Classification Report:
              precision    recall  f1-score   support

           0       0.57      0.55      0.56        49
           1       0.41      0.41      0.41        49
           2       0.29      0.31      0.30        49

    accuracy                           0.42       147
   macro avg       0.43      0.42      0.42       147
weighted avg       0.43      0.42      0.42       147

Confusion Matrix:
[[27  8 14]
 [ 7 20 22]
 [13 21 15]]
```

Models

- Regression and classification models were applied to predict and classify drug-related overall score based on various features such as condition, age, effectiveness, side effects, and benefits..

- Regression models – The Random Forest Regressor performed the best with the lowest RMSE of 1.132 and the highest R^2 score of 0.369, indicating moderately strong predictive performance. Gradient Boosting and XGBoost followed closely with R^2 scores of 0.360 and 0.341, respectively, showing consistent accuracy and good generalization. SVR and SGD Regressor delivered comparatively poor results, with R^2 values close to 0.0, suggesting limited predictive capability. The Decision Tree Regressor also underperformed, with an R^2 of 0.235, possibly due to overfitting. Overall, regression models provided moderate predictive power, with ensemble methods outperforming linear and tree-based single estimators.

- Classification models – Among all classification models, the Random Forest Classifier achieved the highest accuracy of 71.28% and a macro F1-score of 0.497, indicating the best balance between precision and recall across classes. XGBoost and Gradient Boosting Classifier followed, with F1-scores of 0.481 and 0.471, offering good performance especially in detecting higher and lower rating classes. Decision Tree and SVC models had moderate performance but struggled with minority classes. SGD Classifier performed poorly, with a macro F1-score of 0.334, highlighting its limitations in handling class imbalance. Overall, classification models showed good accuracy, with ensemble techniques (Random Forest, XGBoost, Gradient Boosting) clearly outperforming other models, though class imbalance remained a challenge, impacting recall for minority ratings (1, 2, 3).

Final Observations & Storyline

- Acne leads in drug diversity with 39 unique medications (7.98%), followed by Hypertension (34, 6.95%) and Pain (30, 6.13%), showing a wide range of treatment options. Erectile Dysfunction drugs have the highest average overall score (17.90), while Cancer and Stroke drugs score the lowest (2.54 and 3.01), reflecting varied user satisfaction across conditions.

- Bisacodyl stands out with 1,357 reviews, indicating heavy usage, whereas Neomycin has only 1 review, suggesting niche application. The most common side effects—lip/tongue irritation and stomach pain—highlight key patient safety concerns. No significant correlation exists between number of reviews and user ratings (R^2=0.002, p=0.326), indicating ratings aren't driven by review volume.

- Drugs available as both prescription and OTC have the highest average ratings (6.97), slightly better than Rx-only (6.84) and OTC-only (6.69) drugs, suggesting better balance between accessibility and effectiveness. Common generic drugs focus on acne and diabetes, while top brand drugs like Bronkaid and Sudafed dominate respiratory treatments.

- For pregnancy safety, Diabetes drugs lead low-risk categories with 12 drugs each, while Hypertension and Pain dominate high-risk groups (22 and 21 drugs). Alcohol restrictions mostly affect Diabetes and Pain medications, accounting for over 34% of restricted drugs, emphasizing patient caution.

- Statistical tests confirm prescription drugs have significantly higher average scores (8.89 vs. 6.38; t=4.877, p=0), and surprisingly, higher pregnancy risk drugs score better (8.87 vs. 7.44; t=3.017, p=0.0027). Overall, drug ratings reflect a trade-off between efficacy, safety, and accessibility.