Case Study-Drug Efficacy

Team Orion- Prashant Srivastava, Sanya Saxena, Harsh Tantak

Business Objective

Business Case

To evaluate the real-world efficacy of Drug_D and Drug_S by analyzing patient outcomes while accounting for selection bias. The goal is to provide unbiased, actionable insights into which drug is more effective in reducing asthma exacerbations in the first year of treatment.

Scope

- Address the challenges of non-randomized treatment assignment.
- Analyse key pre-treatment variables to adjust for selection bias.
- Compare the post-treatment outcomes of Drug_D and Drug_S fairly and accurately.

Expected Outcome

- A detailed comparison of the efficacy of Drug_D and Drug_S, adjusted for confounding variables.

Dataset Description

This is a real-life dataset used by LifeMed Research which contains asthma patient data, including demographics, comorbidities, prior asthma treatment history, and healthcare costs. It records whether patients were described as Drug_D or Drug_S and tracks asthma exacerbations in the first year post-treatment. This Dataset is divided into different parts-

- Identity Feature
- Pre-Index (before drug assignment)
- Pre-Index Usage (Asthma Treatment History)
- Post-Index (after drug assignment)

Identity Feature-

Pat-Id	Unique patient identifier

Pre-Index (before drug assignment /Previous Diagnostic History)-

Variable Name	**Type of Data**	**Sample Value**	**Description**
Index Age	Quantitative	Range-(14-63)	Age when treatment started
Female	Binary	1-Female 0-Male	Gender of the patient
Pneumonia	Binary	1-Diagnosed with pneumonia last year 0-not Diagnosed with pneumonia last year	Shows that whether the patient was diagnosed with pneumonia last year
Sinusitis	Binary	1- Diagnosed with sinusitis last year 0-not Diagnosed with sinusitis last year	Shows that whether the patient was diagnosed with sinusitis last year

Acute Bronchitis	Binary	1-Diagnosed with Acute Bronchitis last year 0-not Diagnosed with Acute Bronchitis in last year	Shows that whether the patient was diagnosed with Acute Bronchitis last year
Acute Laryngitis	Binary	1-Diagnosed with Acute_ Laryngitis last year 0-not Diagnosed with Acute Laryngitis in last year	Shows that whether the patient was diagnosed with acute Laryngitis last year
Upper_respiratory_infection	Binary	1- Diagnosed with an infection last year 0-not Diagnosed with infection_ last year	Shows that whether the patient was diagnosed with upper respiratory last year
Gerd (Gastroesophageal Reflux Disease)	Binary	1- Diagnosed with Gerd last year 0-not Diagnosed with GERD last year	Shows that whether the patient was diagnosed with Gerd last year
Rhinitis	Binary	1- Diagnosed with Rhinitis last year 0-not Diagnosed with Rhinitis last year	Shows that whether the patient was diagnosed with Gerd last year

Pre-Index Usage (Asthma Treatment History)-

Variable Name	Type of Data	Sample Value	Description
Previous asthma drugs	Quantitative	1 (all the patients in this dataset used 1)	Number of different asthma medications used before this treatment.
Total pre-index canisters (365)	Indicator	Possible values- {0,1,2} Higher value means extreme asthma	Number of Short-Acting Beta Agonist (SABA) canisters used in the previous year.
Pre-Asthma Charge	Quantitative	Minimum Value- 0 Maximum value- 7980	Total asthma related medical charges in last 6 months
Pre-Asthma Days	Quantitative	Minimum Value- 0 Maximum value- 71	Total number of days with asthma treatment in the last year.
Total Pre index Charge	Quantitative	Minimum Value- 1 Maximum value- 875872.5806	Total medical costs in the year before treatment (includes asthma and other conditions).
Drug_S	Indicator	1-Drug_S 0-Drug_D	Type of Drug taken by patient
Pre-Asthma Pharma Charge	Quantitative	Minimum Value- 1 Maximum value- 5463.14	Total asthma-related pharmaceutical charges in the last year.

Post-Index Measurement-

Post Index Exacerbations_365	Quantitative	Range-[0,14]	Number of asthma exacerbations in a year after treatment

Data Preparation

Upon further inspection of the given dataset, we have observed that the data is non-randomized and exhibits selection bias in the whole dataset. Therefore, we have made some assumptions regarding this dataset

- Remove the **previous_asthma_drugs** column from the dataset because every record has the same value, indicating all the patients were already diagnosed with asthma. Additionally, further analysis shows that this variable exhibit very high multicollinearity-
    1. Redundancy- If a variable is highly correlated with others, it does not provide new information and can be removed without affecting the model's predictive power
    2. Unstable Coefficients-In regression models, multicollinearity inflates standard errors, making coefficient estimates unreliable and sensitive to small changes in data
    3. Misinterpretation Of Effects-When predictor variable becomes highly correlated it becomes difficult to determine the individual effect of each variable on the outcome
    4. Overfitting- Increases model complexity which may lead to overfitting reducing the generalizability of the model to the new data

Selection Bias-

Selection bias occurs when the way patients are assigned to treatment groups (Drug_D vs. Drug_S) is **not random**, leading to systematic differences between the groups. In your dataset, selection bias is evident due to **non-random treatment assignment**, which affects the validity of comparing outcomes between Drug_D and Drug_S.

Convenience Bias-

This type of bias occurs when dataset is collected based on ease of access based on access rather than representatives, leading to biased conclusions.

- Only patient with prior asthma treatment are included
- Non-random selection of Drug_D and Drug_S users

Balanced Dataset-

Imbalanced Dataset- A dataset where the distribution of classes or categories is unequal.

Our dataset shows that the distribution of pre-treatment characteristics differs significantly between two treatment groups Drug_s=1 and drug_D=0. We assess this imbalance using the standardized mean difference (SMD)

Which measures the difference in means (for continuous variables) or proportions (for categorical variables) between groups, standardized by a pooled measure of variability

Standardized Mean Difference (SMD)-

- Measures the **difference in means (or proportions)** between the two groups, standardized by the pooled standard deviation.
- It is **not affected by sample size**, making it better than p-values for baseline comparison.

$$SMD = \frac{Mean_{DrugS} - Mean_{DrugD}}{SD_{pooled}}$$

$$SD_{pooled} = \sqrt{\frac{SD_{DrugS}^2 + SD_{DrugD}^2}{2}}$$

Range of SMD

- **SMD < 0.1** → Negligible imbalance (well-matched groups)
- **0.1 ≤ SMD < 0.25** → Moderate imbalance
- **SMD ≥ 0.25** → Large imbalance (significant selection bias)

Standardized Mean Difference (SMD) Results:

Variable	SMD	Interpretation
index_age	0.054800	No Imbalance (SMD < 0.1)
pre_asthma_charge	0.144525	Moderate Imbalance (0.1 ≤ SMD < 0.25)
pre_asthma_days	0.291470	Large Imbalance (SMD ≥ 0.25)
log_charges	0.050608	No Imbalance (SMD < 0.1)
log_asthma_charge	0.646011	Large Imbalance (SMD ≥ 0.25)
female	0.041084	No Imbalance (SMD < 0.1)
pneumonia	0.079905	No Imbalance (SMD < 0.1)
sinusitis	0.011408	No Imbalance (SMD < 0.1)
acute_bronchitis	0.149202	Moderate Imbalance (0.1 ≤ SMD < 0.25)
acute_laryngitis	0.028707	No Imbalance (SMD < 0.1)
gerd	0.091678	No Imbalance (SMD < 0.1)
rhinitis	0.352552	Large Imbalance (SMD ≥ 0.25)

Interpretation	Count
No Imbalance (SMD < 0.1)	7
Large Imbalance (SMD ≥ 0.25)	3
Moderate Imbalance (0.1 ≤ SMD < 0.25)	2

This shows that this dataset shows large imbalance of data

To balance the data, we use IPTW (Inverse Probability Treatment Weighting)

IPTW- (Inverse Probability Treatment weighting)-

It is a statistical technique used to adjust for selection bias in observational Studies. It creates a pseudo-randomized scenario where the distribution of covariates is balanced between treatment groups

Propensity Score- The propensity score is the probability of receiving Drug_S (drug_s = 1) given the observed covariates. Calculated using logistic regression

$$P(T=1|X) = \frac{e^{\beta_0+\beta_1 X_1+\beta_2 X_2+\ldots+\beta_n X_n}}{1+e^{\beta_0+\beta_1 X_1+\beta_2 X_2+\ldots+\beta_n X_n}}$$

Calculating Inverse Probability Weights

- $P(T=1|X_i)$ propensity score for patient i
- If the patient received Drug_S (T=1), their weight is $\frac{1}{P(T=1|X_i)}$
- If the patient received Drug_D (T=0) their weight is $\frac{1}{1-P(T=1|X_i)}$

Using IPW for adjustment

After obtaining IPW Weights, they are used to re-weight the samples such that both treatment groups become comparable

Adjusted Outcome was given by

$$E[Y|T] = \frac{\sum_i Y_i \cdot W_i}{\sum_i W_i}$$

We ensure that the covariate distributions in the treated and untreated groups are similar

Here is the updated column description of the data

Propensity Score	Probability	Probability that a patient recives Drug_S(drug_S=1)
IPW Weight	Quantitative	Adjust seletion bias by creating a pseudo_randmoized sample

X-Axis (Trimmed IPW)

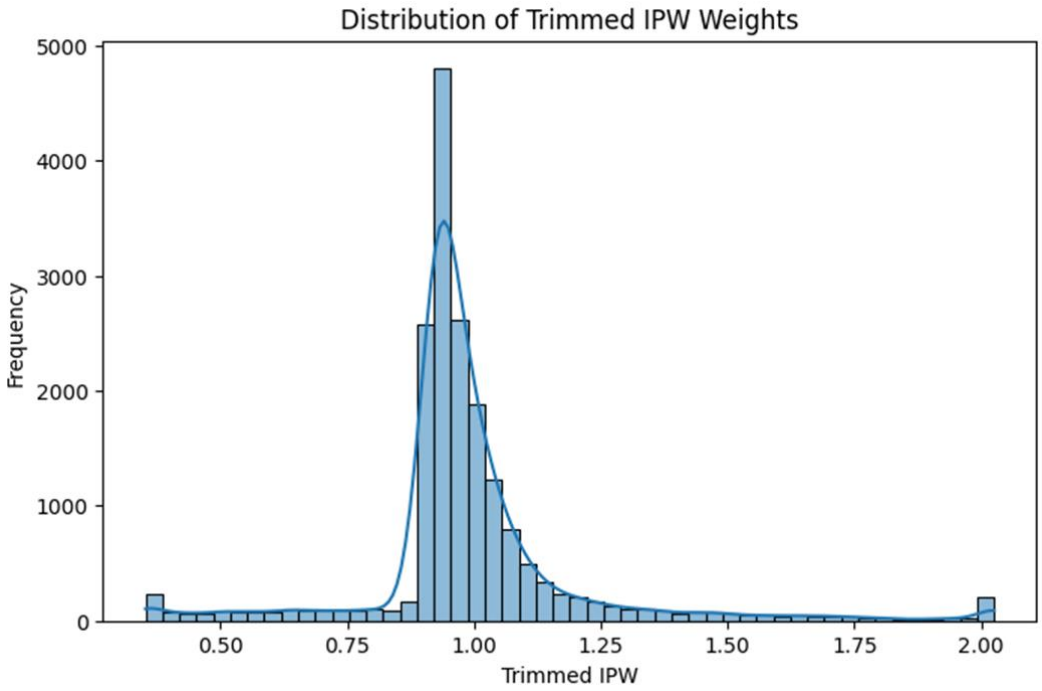
- Represents the **Inverse Probability Weights (IPW)** after trimming.
- IPW is commonly used in causal inference to balance treatment groups by reweighting observations.

Y-Axis (Frequency)

- Represents the **number of observations** for each weight value.

Histogram with Kernel Density Estimate (KDE) Curve

- The histogram shows the frequency distribution of IPW values.
- The KDE curve provides a smoothed representation of the probability density.



Key Findings

1. **Most Weights Centered Around 1**
    - The majority of the weights are **close to 1**, suggesting that most observations have **similar importance** in the weighted analysis.

- This is a good sign because extreme weights can lead to instability in statistical models.

2. **Right Skewed Distribution**
    - The distribution has a long **right tail**, indicating some observations have **higher weights**.
    - This suggests that a few cases contribute more to the weighted model, potentially due to **underrepresented groups**.
3. **Trimming Effect**
    - The presence of a sharp cutoff on the left and right suggests **trimming** has been applied to remove extreme weights.
    - This helps prevent **large variance issues** and makes the model **more stable**.
4. **Some Outliers at the Extremes**
    - A few observations with **low (around 0.5) and high (near 2.0) weights** remain.
    - These outliers may represent cases that are **hard to balance** and may require further investigation.

Conclusion

- The IPW weights are **well-distributed** around 1, indicating good balance in the dataset.
- The **right skew** suggests some observations receive more weight, but trimming has controlled extreme values.
- Further analysis could examine whether these high-weight cases impact the final model.

By using Principal Component Analysis we have categorized severity of asthama patients

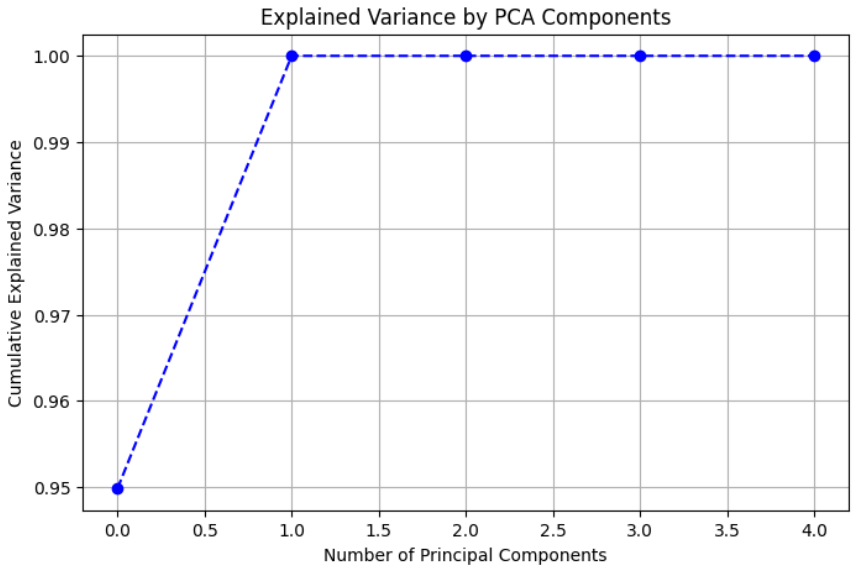
Severity Score	Quantitative	Minimum Value-0 Maximum Value-100	Score representing severity of a patient condition Higher score likes indicates more severe of a patient condition
Severity Category	Categorical	Mild-0-33 Moderate-34-66 Severe-67-100	Classify patients as mild,moderate or severe

The selected features (previous_asthma_drugs, total_pre_index_cannisters_365, pre_asthma_days, pre_asthma_charge, pre_asthma_pharma_charge) are standardized using Standard Scalar (), ensuring that all features have a mean of 0 and a standard deviation of 1

PCA reduces the dimensionality of the standardized data while preserving the most important variance. Here, the top 3 principal components (PCs) are selected.

The severity score is computed as a weighted sum of the top 3 principal components, where weights are the explained variance ratios of each component.

The severity score is transformed into a 0-100 range for easier interpretation.



Axis-

X-axis- Number of Principal components

Y-axis-Cumulative Explained Variance

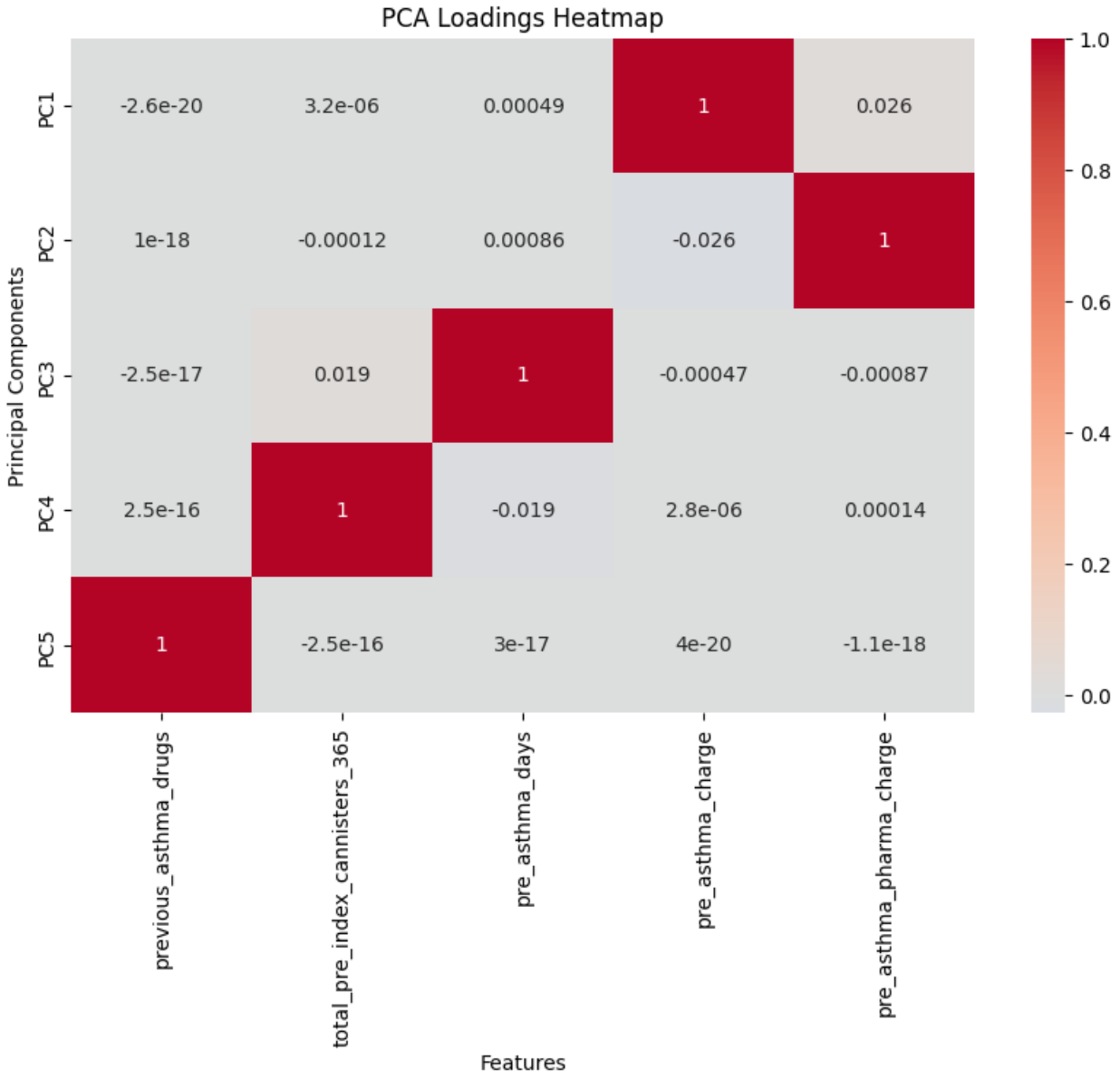
Interpretation-

- The first principal component (PC1) explains approximately **95%** of the variance
- The second principal component (PC2) increases the cumulative explained variance to **100%**, meaning that together, the first two components fully capture the variance in the dataset.
- Adding more components (PC3, PC4, etc.) does not contribute further variance, as seen in the flat line at 100%.
- The steep rise after PC1 shows that most of the dataset's variance is captured by the first component.
- Since **only two components** explain nearly **100%** of the variance, reducing the dataset to just these two dimensions would likely retain most of the information while simplifying the model.

PCA Loadings Heatmap-

Heatmap visualizes the PCA loadings, which indicate how strongly each original feature contributes to the principal components (PCs). The values in

the heatmap represent the correlation (or weight) of each feature with each principal component.



Axes And Labels-

X-Axis (Features)- Original dataset variables

- previous_asthma_drugs
- total_pre_index_canisters_365
- pre_asthma_days
- pre_asthma_change

- pre_asthma_pharma_change

Y-Axis (Principal Components - PCs): The transformed components from PCA, are labeled as **PC1, PC2, PC3, PC4, and PC5**.

Interpreting the Heatmap

- The **color scale** represents the magnitude of the loading values:
    - **Dark red (1 or -1)**: Strong contribution of the feature to that principal component.
    - **Lighter shades/grey**: Low or near-zero contribution.

Key Observations:

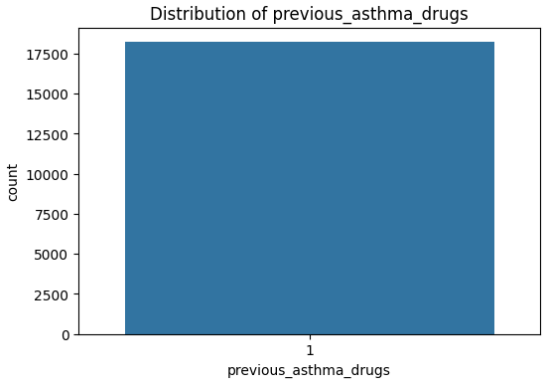
- pre_asthma_change has the strongest positive loading for **PC1 (value ≈ 1)**, meaning it contributes the most to this component.
- pre_asthma_pharma_change strongly contributes to **PC2**.
- pre_asthma_days is the dominant feature in **PC3**.
- total_pre_index_canisters_365 is the strongest in **PC4**.
- previous_asthma_drugs is the main feature for **PC5**.
- Many other values are very close to zero (light grey), meaning those features have little or no impact on the respective principal component.

Exploratory Data Analysis

Class Distribution of Categorical Variables

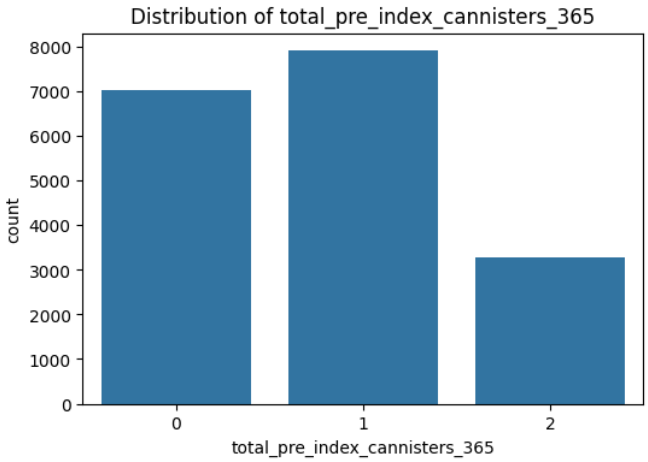
Previous Asthma Drugs:

- The distribution is entirely concentrated at 1, suggesting that all individuals in the dataset had used asthma medication before.



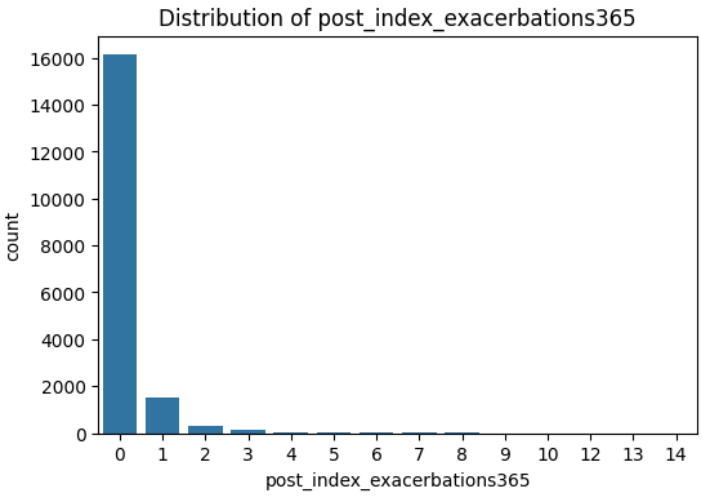
Total Pre-Index Canisters (365 Days):

- Shows three categories (0, 1, 2).
- Most people had either 0 or 1 canister usage in the past year, with fewer at 2.



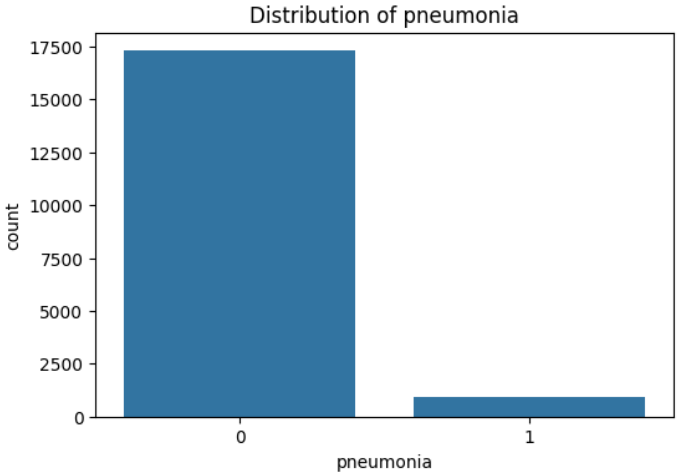
Post-Index Exacerbations (365 Days):

- Highly skewed distribution with most individuals having 0 exacerbations, and a small proportion having 1 or more.
- The frequency drops significantly as the number of exacerbations increases.



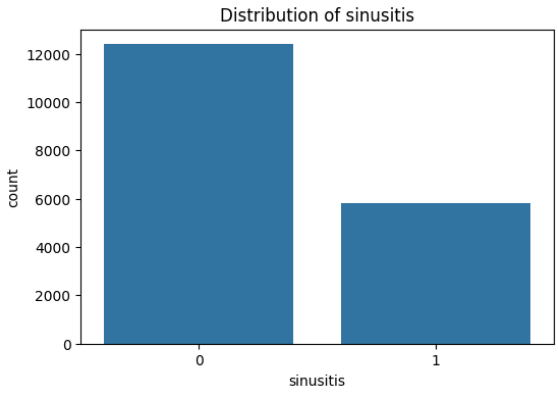
Pneumonia:

- The majority (about 90%) do not have pneumonia, while a small fraction does.



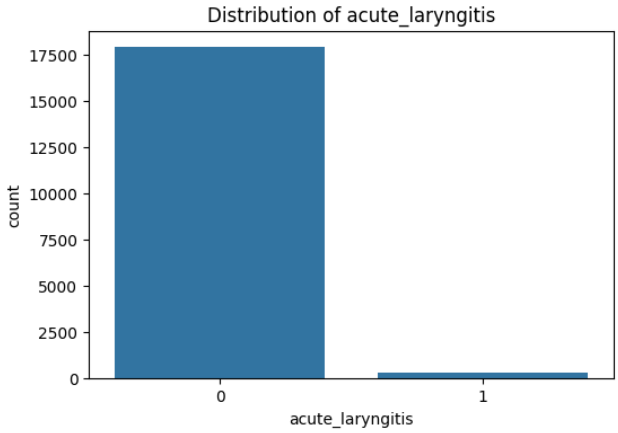
Sinusitis:

- The presence of sinusitis is more common than pneumonia but still occurs in a minority of cases.



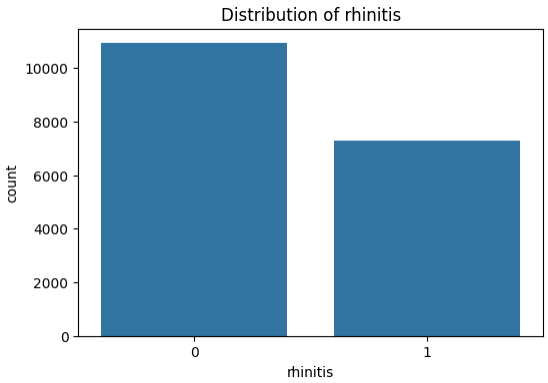
Acute Laryngitis:

- Very few cases of acute laryngitis exist in the dataset.



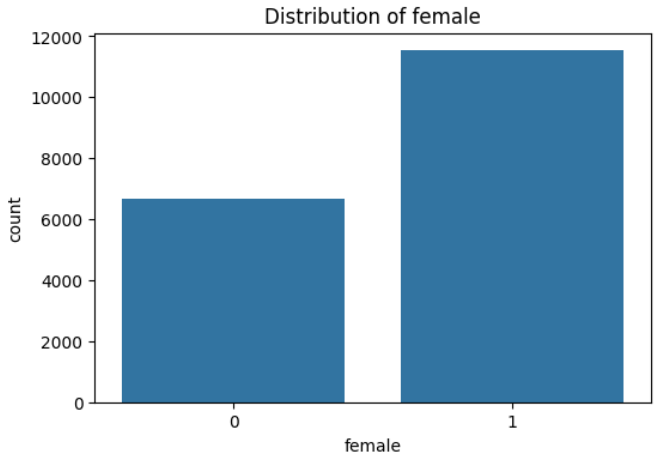
Rhinitis:

- More evenly distributed compared to the previous conditions, with a significant portion affected.



Gender (Female = 1, Male = 0):

- More females than males in the dataset.



General Insights:

- The dataset contains a majority of individuals who have previously used asthma drugs.
- Most individuals have low or no prior canister usage.
- Severe asthma exacerbations are rare.
- Comorbid conditions like pneumonia and laryngitis are uncommon, whereas sinusitis and bronchitis are relatively more prevalent.
- There are more females than males in the dataset.

Statistical Tests of Association-

Chi-Square Test Results

Condition	Chi2 Statistic	p-value	Significant Association?
Acute Bronchitis	36.0001	0.000593	Yes
Acute Laryngitis	71.1558	4.91e-10	Yes
Upper Respiratory Infection	23.5142	0.0359	Yes
GERD	32.8776	0.00178	Yes
Rhinitis	23.1292	0.0402	Yes
Drug S	22.1555	0.0530	No

t-Test Results

Variable	**t-Statistic**	**p-value**	**Significant Difference?**
Index Age	-0.5173	0.6050	No (Fail to Reject H_0)
Total Pre-Index Cannisters (365)	-0.9165	0.3594	No (Fail to Reject H_0)
Total Pre-Index Charge	-3.3840	0.000716	Yes (Reject H_0)
Pre-Asthma Days	3.3143	0.000920	Yes (Reject H_0)
Pre-Asthma Charge	-0.6339	0.5262	No (Fail to Reject H_0)
Pre-Asthma Pharma Charge	2.2522	0.0243	Yes (Reject H_0)
Log Charges	-3.9356	8.33e-05	Yes (Reject H_0)
Log Asthma Charge	3.0859	0.00203	Yes (Reject H_0)

Model Performance

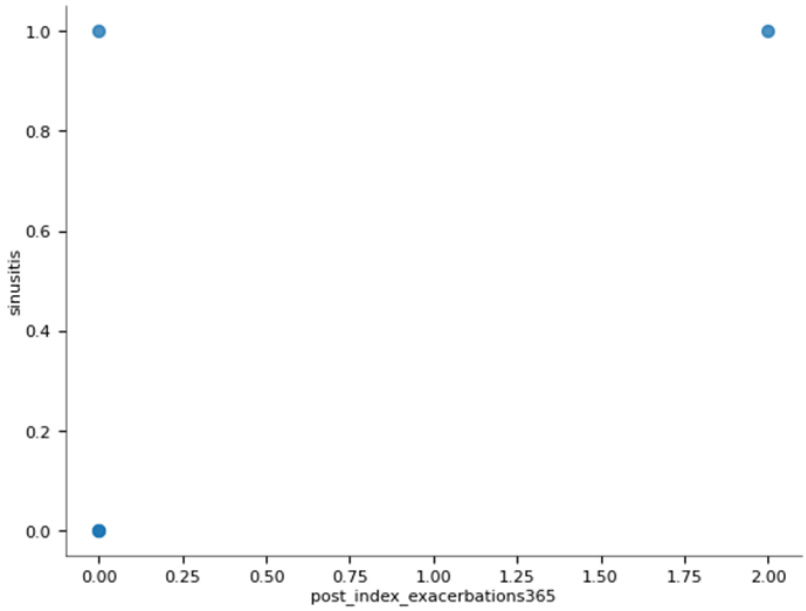
Working of the XG-Boost Model for Predicting Exacerbations

The XG-Boost model in this instance is used for predicting the number of exacerbations a patient is likely to undergo within a specified timeframe. The dataset is preprocessed first by choosing important features and discarding redundant variables such as propensity scores and prior history with asthma drugs to prevent data leakage.
Some numerical features based on costs and severity scores are standardized so that the model treats them equally. Also, Inverse Probability Weighting (IPW) is used to correct **confounding variables, generating a more representative dataset**.
The data are then divided into training (80%) and test (20%) sets **to** effectively assess the model's performance.
For training, an XG-Boost Regressor with a Poisson objective function (count: Poisson) is employed, which is suitable for making predictions based on counts such as exacerbation frequency. The model is set to optimize the Root Mean Squared Error (RMSE), a standard measure for regression tasks.
IPW weights are used in training to make the model capture real-world patterns. XG-Boost, which excels at capturing intricate relationships, learns interactions among several variables in the dataset efficiently.
Once trained, the model makes predictions of exacerbation counts for the test dataset. These forecasts provide an approximation of the number of exacerbation episodes a patient is predicted to have according to previous medical history and corresponding cost or severity measures.
Utilizing gradient boosting, the model repeatedly adjusts its forecasts so that it becomes extremely efficient at identifying underlying patterns in the data.
The primary benefit of employing XG-Boost in this scenario is that it can model non-linear relationships and interactions in structured tabular data. It is also sturdy in dealing with missing values and can efficiently accommodate IPW weighting, which is important in minimizing biases in observational studies. Though the model necessitates sensitive hyperparameter tuning to perform optimally, it yields a good trade-off between accuracy and flexibility and thus qualifies as a viable option for such a predictive task.

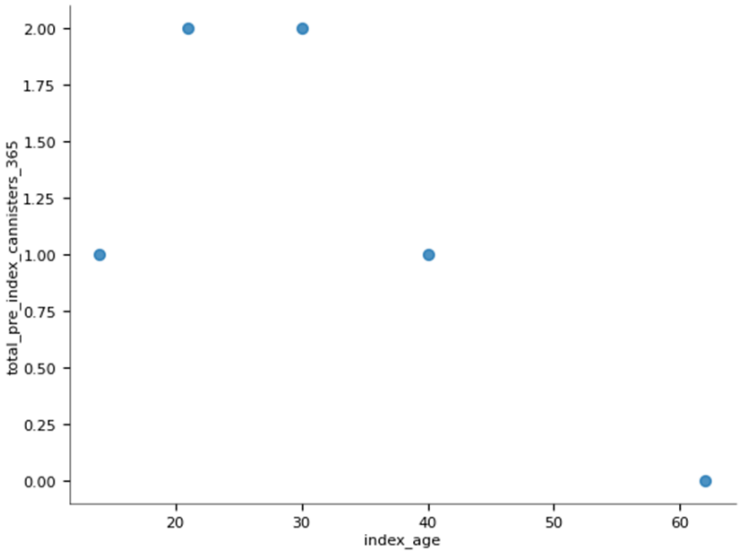
Model Accuracy-

The model was successful, with an RMSE of 0.4521 and an MAE of 0.2154, which means its predictions are close to actual counts of exacerbations. The R^2 score of 0.6825 also implies that the model is highly effective at explaining the variance of the data and is a top pick for this predictive task.

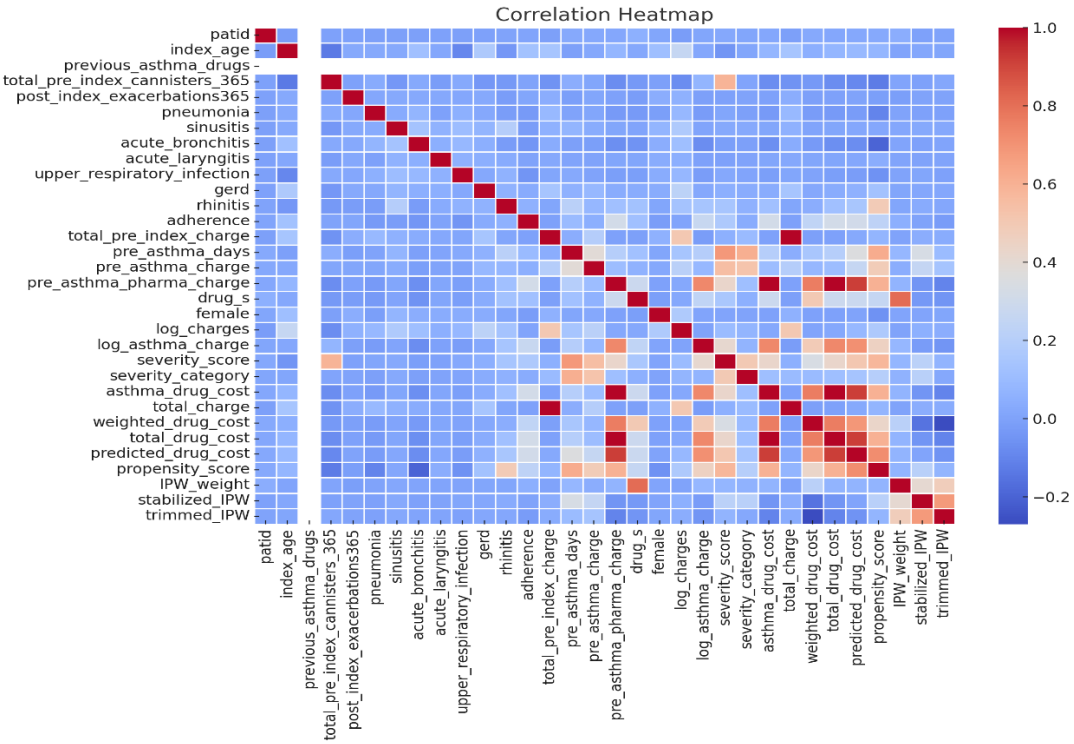
- X-axis: post_index_exacerbations365 (number of exacerbations in the past 365 days after an index event)
- Y-axis: sinusitis (binary variable, 0 or 1, indicating whether sinusitis is present)
- The data points suggest sinusitis occurs only at specific values of exacerbations, indicating a possible categorical or sparse dataset.



- X-axis: index age (age at the index event)
- Y-axis: total_pre_index_cannisters_365 (number of canisters used in the past year before the index event)
- The scattered points indicate the usage of canisters is not uniform across ages, suggesting possible clusters or specific age groups with higher usage.



1. **Strong Positive Correlations (Closer to +1)**
    - If two variables **correlate close to +1**, it means they increase together.
    - Example: If "total_pre_index_cannisters_365" and "post_index_exacerbations365" show a strong positive correlation, it suggests that patients who used more inhaler cannisters before the index event had more exacerbations afterward.
2. **Strong Negative Correlations (Closer to -1)**
    - If two variables **correlate close to -1**, it means that when one increases, the other decreases.
    - Example: If "index_age" and "total_pre_index_cannisters_365" show a negative correlation, it suggests that older patients used fewer inhaler cannisters before the index event.
3. **Weak or No Correlation (Close to 0)**
    - If the correlation is **close to 0**, there is no strong linear relationship.
    - Example: If "sinusitis" and "index_age" have a near-zero correlation, it means there is no clear relationship between sinusitis and age.
4. **Target Variable (Severity Category) Correlations**
    - If "severity_category" is included, we check which features have the highest correlation with it.
    - Example: If "post_index_exacerbations365" has a high correlation with severity_category, it suggests that more exacerbations after the index event are linked to severe cases.


Correlation Heatmap
patid
index_age
previous_asthma_drugs
total_pre_index_cannisters_365
post_index_exacerbations365
pneumonia
sinusitis
acute_bronchitis
acute_laryngitis
upper_respiratory_infection
gerd
rhinitis
adherence
total_pre_index_charge
pre_asthma_days
pre_asthma_charge
pre_asthma_pharma_charge
drug_s
female
log_charges
log_asthma_charge
severity_score
severity_category
asthma_drug_cost
total_charge
weighted_drug_cost
total_drug_cost
predicted_drug_cost
propensity_score
IPW_weight
stabilized_IPW
trimmed_IPW
1.0
0.8
0.6
0.4
0.2
0.0
−0.2

Compare the post-treatment outcomes of Drug_D and Drug_S fairly and accurately.

We have calculated some different statistical metrics to compare drugS and drugD

- Mean Predicted Exacerbations-

  Drug_S=0.0800

  Drug_D=0.1648

  Interpretation:

  - Patients on **Drug_S** experience significantly fewer exacerbations on average compared to those on **Drug_D**.
  - This suggests that **Drug_S is more effective in preventing exacerbations** in asthma patients.

- Statistical Significance Test-

  T-Statistic: -20.2448

  P-value: 0.0000 (highly significant)

  Conclusion:

  The difference in exacerbations between Drug_S and Drug_D is statistically significant (not due to random variation).

  Interpretation:

  - The extremely low **p-value (< 0.05)** indicates that the difference between Drug_S and Drug_D is **not due to random variation**.
  - A t-statistic of **-20.2448** (a large negative value) confirms that **Drug_S has a significantly lower mean exacerbation rate than Drug_D**.
  - This means we can confidently say that **Drug_S is superior in reducing asthma exacerbations**.

- Risk Reduction

  Absolute Risk Reduction (ARR)

  ARR = Mean exacerbations in Drug_D – Mean exacerbations in Drug_S

ARR = 0.1648 - 0.0800 = 0.0768

Patients on Drug_S experience 0.0768 fewer exacerbations per patient compared to Drug_D.

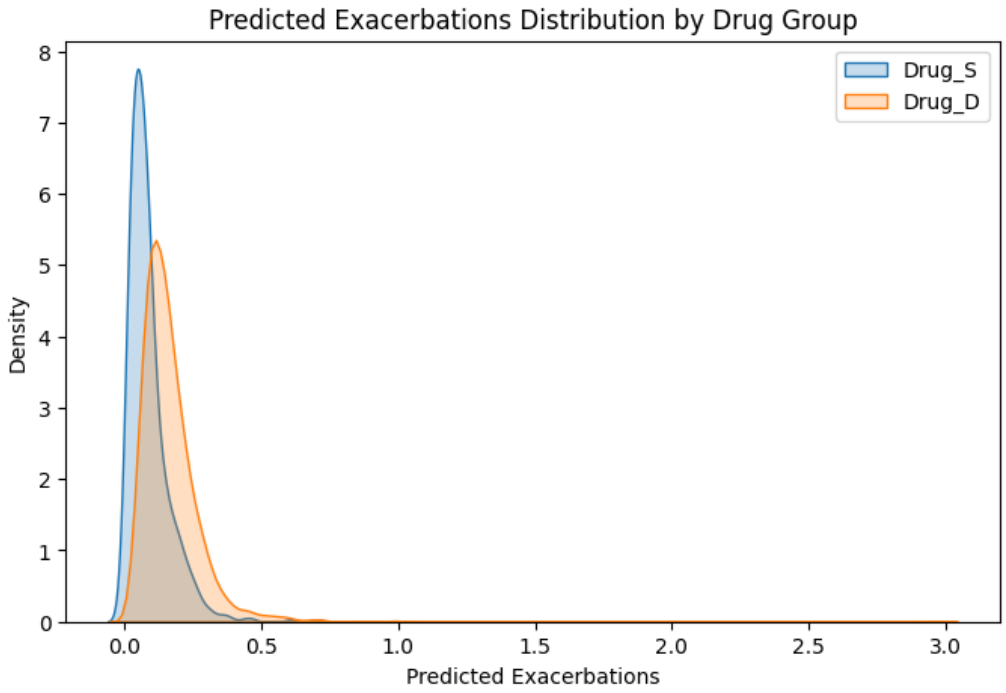
Interpretation:

- A **47% risk reduction** is **clinically significant**, meaning Drug_S has a strong protective effect against exacerbations.
- A high RRR suggests that **switching from Drug_D to Drug_S could substantially benefit asthma patients** by reducing their risk of exacerbations.

KDE (kernel Density Estimation) Plot-

A KDE plot was used to visualize the distribution of exacerbation rates for each drug:

- **Drug_S (blue curve)**:
  - The curve is **shifted left**, indicating that most patients experience **fewer exacerbations**.
  - The **distribution is more concentrated**, meaning Drug_S provides **consistent and predictable** results.
- **Drug_D (orange curve)**:
  - The curve is **shifted right**, showing a higher number of exacerbations.
  - The **distribution is more spread out**, suggesting that Drug_D leads to **greater variability** in patient outcomes.



Interpretation:

- Drug_S is **not only better on average but also more reliable** in reducing exacerbations.
- Drug_D leads to a **higher and more variable risk** of exacerbations.

Final Conclusion

Based on the evidence from statistical analysis and visualization:

- **Drug_S is significantly better at reducing exacerbations** than Drug_D (**p-value = 0.0000**).
- **The risk reduction is nearly 47%**, which is **clinically meaningful** and suggests a strong benefit of Drug_S.
- **The KDE plot confirms that Drug_S leads to fewer and more consistent exacerbations**, while Drug_D results in more frequent and variable exacerbations.

Clinical Recommendation-

Drug_S should be the preferred choice for reducing exacerbations in asthma patients. Given Statistically significant and clinically meaningful benefits, it is more likely the more effective safer for managing asthma

Market Analysis (Drug-S Vs Drug-D)

Drug_D has lower costs overall –

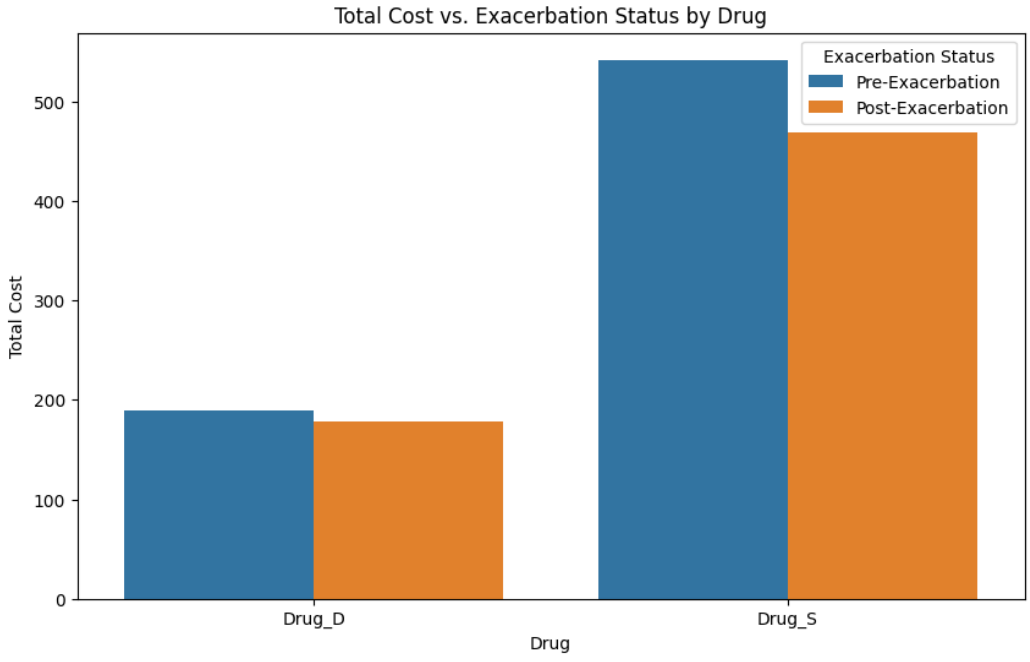
- Both pre-exacerbation and post-exacerbation costs for **Drug_D are significantly lower** than Drug_S.
- Post-exacerbation costs **slightly decrease**, indicating **lower financial impact post-treatment**.

Drug_S has a higher cost burden –

- The total cost for **Drug_S is significantly higher** than Drug_D.
- Post-exacerbation costs decrease slightly but remain **much higher than Drug_D's costs**.

Cost Reduction after Exacerbation in Both Drugs –

- Both drugs show **a decrease in total cost post-exacerbation**, but **Drug_S remains much more expensive**.
- **Drug_D might be more cost-effective** – Since it has **lower pre- and post-exacerbation costs**, it may be a more **economical treatment option**.
- **Drug_S may indicate higher severity patients** – The **higher cost of Drug_S** might suggest it is used for **more severe cases** or requires **additional medical resources**.
- **Effectiveness vs. Cost Trade-off** – If Drug_S leads to **better long-term outcomes**, its higher cost might be justified. However, if Drug_D achieves **similar or better health outcomes at a lower cost**, it may be the preferable option.



Key findings:

1. From the post index excerbations, adherence , severity score drug_d and drug_s respectively

Post-index exacerbations - Drug S: 0.15, Drug D: 0.18

Adherence - Drug S: 0.31, Drug D: 0.24

Severity Score - Drug S: 8.48, Drug D: 6.18

We can evaluate that Drug s has a slightly lower exacerbation rate which suggests that the drug s is more effective at preventing exacerbations. Similarly, for the adherence rate, the patients taking drug s are more likely to have better adherence which will lead to better treatment outcomes. The severity score of s is a bit lower which indicates the opposite result. As a cumulative of result, we may conclude that from this calculation drug s is better than drug d in terms of exacerbation rated.

	severity category	avg_pre_asthma_cost	total_pre_asthma_cost
0	Mild	242.832075	4383118.96
1	Moderate	916.468551	63236.33
2	Severe	1619.633750	12957.07

From the above calculations, we may conclude that cost increases with severity. Here mild cases drive total spending because they are more common. Here moderate cases have a significantly high drug cost. This is because there were fewer patients in the severe category.

Market Segmentation & Drug Positioning:

- Mild Cases Dominate (Mass Market) 18,050 patients (≈99.5%) → This is the largest potential market. Low per-patient cost (242), but highest total cost (4.38M).
- Drug Market Strategy: Competitive low-cost treatment (cheaper than existing options). Potential for preventive strategies to avoid escalation to moderate/severe stages.
  - Moderate Cases (Premium Market) Only 69 patients (~0.4%) → Small segment, but spending is high (916 per patient). Total cost is low(63K), but per-patient cost is 3.7x higher than mild cases.
  - Drug Market Strategy: ✓ Offer mid-tier pricing—moderate patients are likely willing to pay for better effectiveness.
  - ✓ Could be a target for insurance reimbursements if the drug shows cost savings.

- Severe Cases (Niche, High-Cost Segment) Only 8 patients (~0.04%), but each spends 1,619 on pre−asthma treatments. The total cost is only 12,957, so this is a small but expensive-to-treat group. Drug Market Strategy:
  ✓ If the drug significantly reduces exacerbations, it could be a high-cost specialty drug.
  ✓ Target insurers & hospitals for specialized treatment plans.

Weighted Effectiveness Score (Drug S): 1.0114

Weighted Effectiveness Score (Drug D): 0.2350

Drug S is more effective based on adjusted health outcomes.

With a significantly higher Weighted Effectiveness Score (1.0114 vs. 0.2350), Drug S is more effective. Better condition management is indicated by fewer Drug S exacerbations (0.15 vs. 0.18). Drug S's improved performance may be due to higher adherence (0.31 versus 0.24). Drug S patients have a higher severity (8.48 vs. 6.18), meaning that it is effective even in high-risk patients.

Conclusion

The analysis suggests that **Drug S is more effective** than **Drug D**, despite its higher cost. **Drug S has a lower exacerbation rate (0.15 vs. 0.18), higher adherence (0.31 vs. 0.24), and a greater effectiveness score (1.0114 vs. 0.2350)**, indicating better treatment outcomes. While **Drug D is more cost-effective**, the **higher severity score for Drug S (8.48 vs. 6.18)** implies that it is used in more severe cases and still performs well.

From a **market strategy perspective**, **mild cases dominate the patient base (99.5%)**, making cost-effective treatments crucial. However, **moderate cases incur higher per-patient costs**, and **severe cases, though rare, require significant spending**. Thus, a **tiered pricing strategy** could optimize accessibility and profitability. **Insurance partnerships for moderate and severe cases** could further enhance Drug S's market positioning.

Ultimately, if **Drug S provides long-term health benefits**, its **higher cost may be justified**. However, **if Drug D delivers comparable outcomes at a lower cost**, it could be a viable alternative for mild-to-moderate patients.