# Team Ranger CoverMyMeds Project

By: Andy Huynh, Alexander Lago,
Sergei Nenanov, and Zheyu Ni

covermymeds
empowering patients

# Introduction

- This is a project containing artificial pharmaceutical data provided by CoverMyMeds.
- We want to:
  a. provide the costs of medication at a pharmacy
  b. provide the formulary status of the medication
  c. develop a method to group similar medications together
- The data includes date, pharmacy, diagnosis, drug, and three indicators of insurance.
- We also know if insurance rejected the drug, and if not, how much the copay is

| tx_date | pharmacy | diagnosis | drug | bin | pcn | group | rejected | patient_pay |
|---|---|---|---|---|---|---|---|---|
| 2022-01-02 | Pharmacy #6 | G99.93 | branded tanoclolol | 725700 | 1UQC | NaN | False | 13.39 |
| 2022-01-02 | Pharmacy #42 | U60.52 | branded oxasoted | 664344 | NaN | 52H8KH0F83K | False | 7.02 |
| 2022-01-02 | Pharmacy #37 | Q85.91 | branded cupitelol | 725700 | 1UQC | NaN | False | 13.39 |
| 2022-01-02 | Pharmacy #30 | U60.52 | generic oxasoted | 571569 | KB38N | 6BYJBW | False | 10.84 |
| 2022-01-02 | Pharmacy #18 | N55.01 | branded mamate | 664344 | NaN | ZX2QUWR | False | 47.00 |

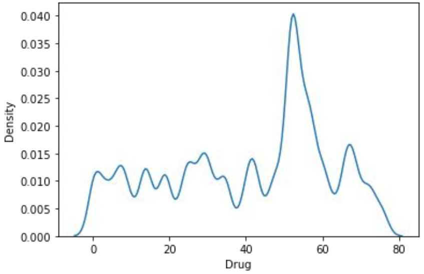
# Exploratory Data Analysis (EDA)

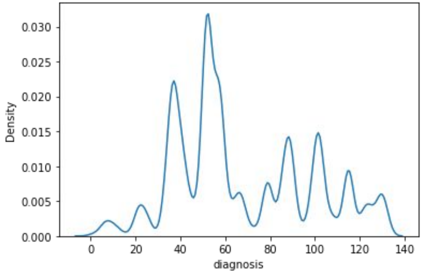
### Rejected on Branded/Generic

- A main focus was the effect of various variables on rejection of a drug by the insurance company.
- We suspected that a drug being branded versus generic would affect rejection rates. To confirm this, we ran a logit regression of rejection on branded. With 0.921 accuracy our regression told us that branded significantly affected rejection by 30%.

### Drug and Diagnosis

- First off, we found 77 categories of Drugs and 133 categories of Diagnosis.
- We have a negatively skewed distribution for Drugs and a positively skewed distribution for Diagnosis indicating that they may affect copay. We later confirm this.



# Exploratory Data Analysis Continued…

**Drug and Diagnosis on Copay**

- In a preliminary linear regression, we found Drug and Diagnosis to affect copay by -6% and 7.7% respectively

**Drug and Diagnosis on Rejected**

- When tested for an effect on rejection Drug and Diagnosis did not have a statistically significant effect on rejection

# Multivariable Regression Result

- After controlling for different drugs, diagnosis and insurance plan,

  branded drug >$3 >generic drug.

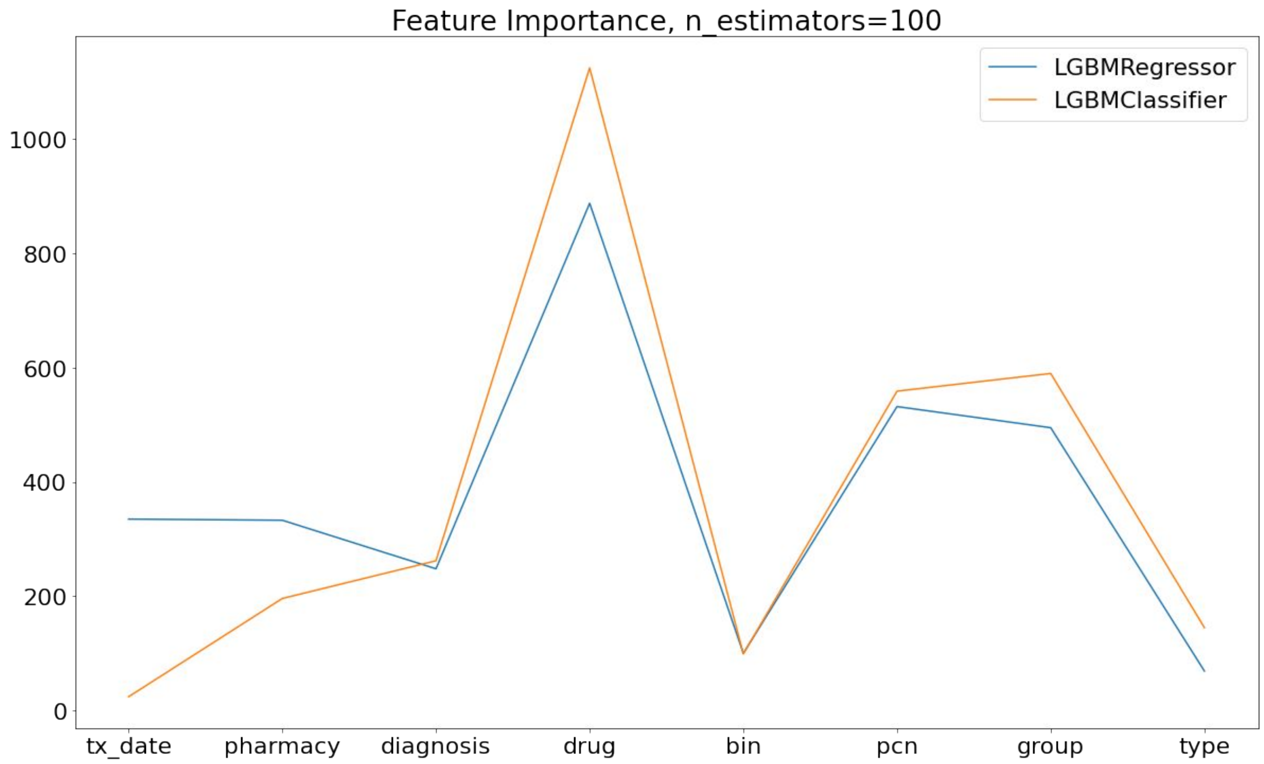
- Most important features: Different drugs, its branded/generic status, insurance plan
- Pharmacy location doesn't have a significant impact on the patient pay
- Prediction power in Rejection Rate: TPR = 0, TNR = 100%. Accuracy rate in the test sample is 92.19%.
- Goodness of fit of regression on patient pay: MSE is 826.64.

# Comparing Different Models

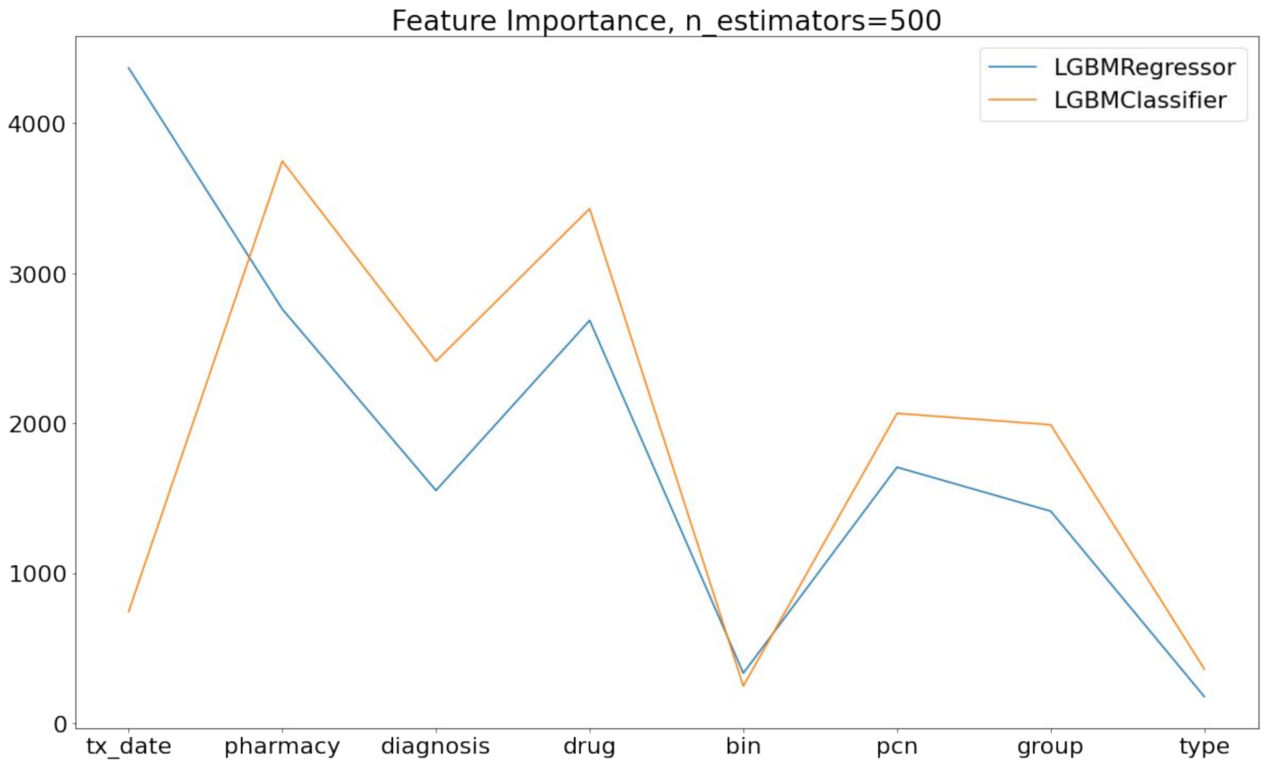
- We establish a baseline with linear regression
- We considered Random Forest but it is too slow
- We used the two modern gradient boosting algorithms
    - LightGBM
    - CatBoost

| | Linear Regression | LightGBM | CatBoost |
|---|---|---|---|
| MSE | 827 | 624 | 640 |

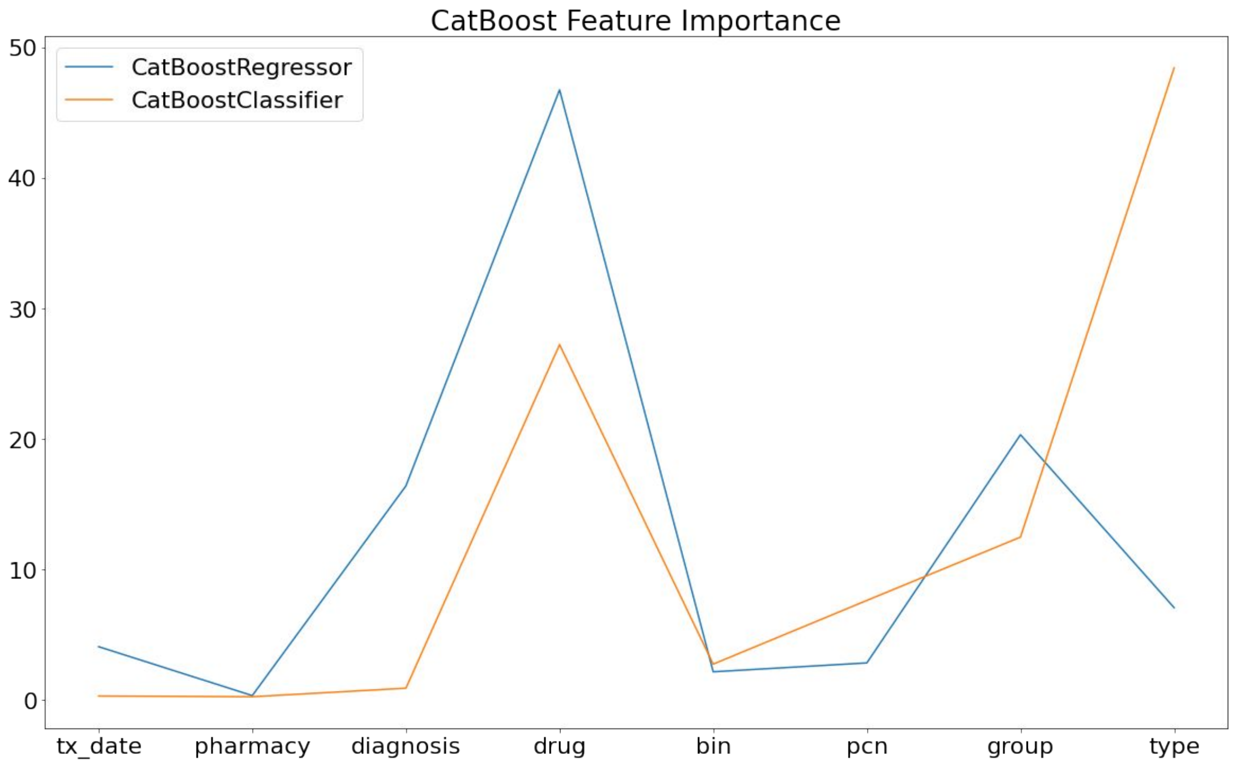
# LightGBM, with n_estimators=100



# LightGBM with n_estimators=500

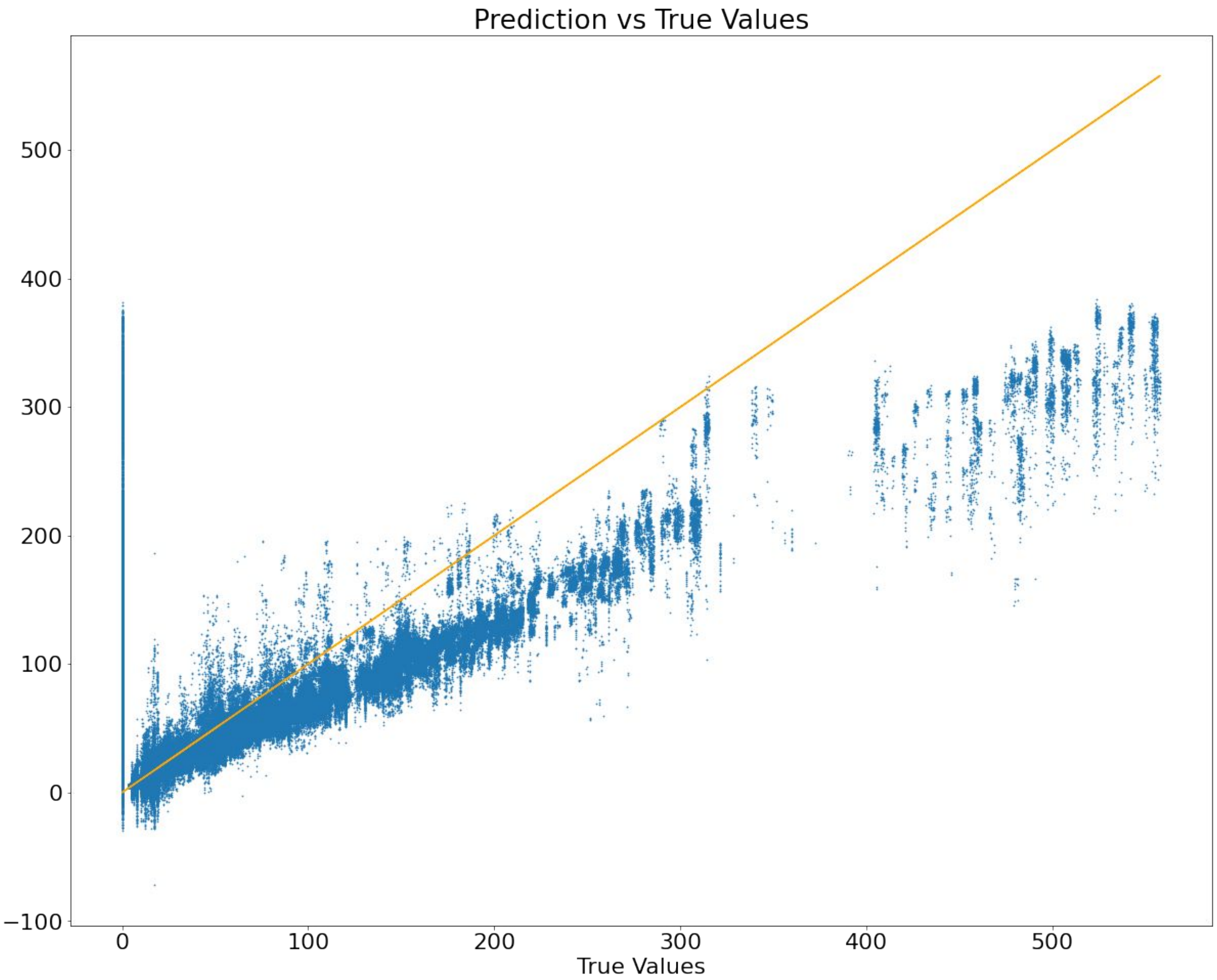


# CatBoost


CatBoost Feature Importance
CatBoostRegressor
CatBoostClassifier
0
10
20
30
40
50
tx_date
pharmacy
diagnosis
drug
bin
pcn
group
type

# LGBM vs CatBoost

- Compared to LGBM, due to how CatBoost ranks feature importances, and how it matches our expectations from our EDA, we have decided to use CatBoost.


Prediction vs True Values
CatBoost Predictions
True Values
500
400
300
200
100
0
−100
0
100
200
300
400
500

# Formulary Status

- Formulary status is a ranking used by insurance companies to determine preferences in paying for treatments for given diagnoses
- Insurance companies are more likely to approve:
  - Inexpensive drugs
  - Common drugs over novel treatments
- To model each insurance plan's formulary status, we used the following to build a model:
  - The patient price to estimate overall price
  - The rejection rate
  - The quantity of drugs requested

# Utilizing MCDM to estimate formulary status

- We used the TOPSIS model of MCDM to favor the cheapest, lowest rejection and most prescribed drug
  - This is how we interpret formulary status.
- The model finds the drug that satisfies this criteria both the best and the worst.
- Then it interpolates and ranks every combination in between the best and the worst.

| Diagnosis | Bin | PCN | Group | Drug | Median Patient Pay | Rejection Rate | Total Count | Absolute Rank | Grouped Ranking |
|---|---|---|---|---|---|---|---|---|---|
| I68.27 | 725700 | 327CKV | IOEAN1DWVV3Y | hidizuzunib | 93.75 | 0.0 | 30956 | 0.393125 | 3.0 |
| I68.27 | 725700 | 327CKV | IOEAN1DWVV3Y | mule | 40.18 | 0.0 | 35231 | 0.423156 | 2.0 |
| I68.27 | 725700 | 327CKV | IOEAN1DWVV3Y | prazinib | 10.62 | 0.0 | 105066 | 0.997551 | 1.0 |

# Thank you for watching!

## We also would like to thank:

- Melanie Butler for her invaluable guidance with this project
- CoverMyMeds for giving us an opportunity to produce practical work for them, with a robust dataset
- Matt Osborne for his detailed lectures, we couldn’t have made this work without them