



# Drug Review Analysis

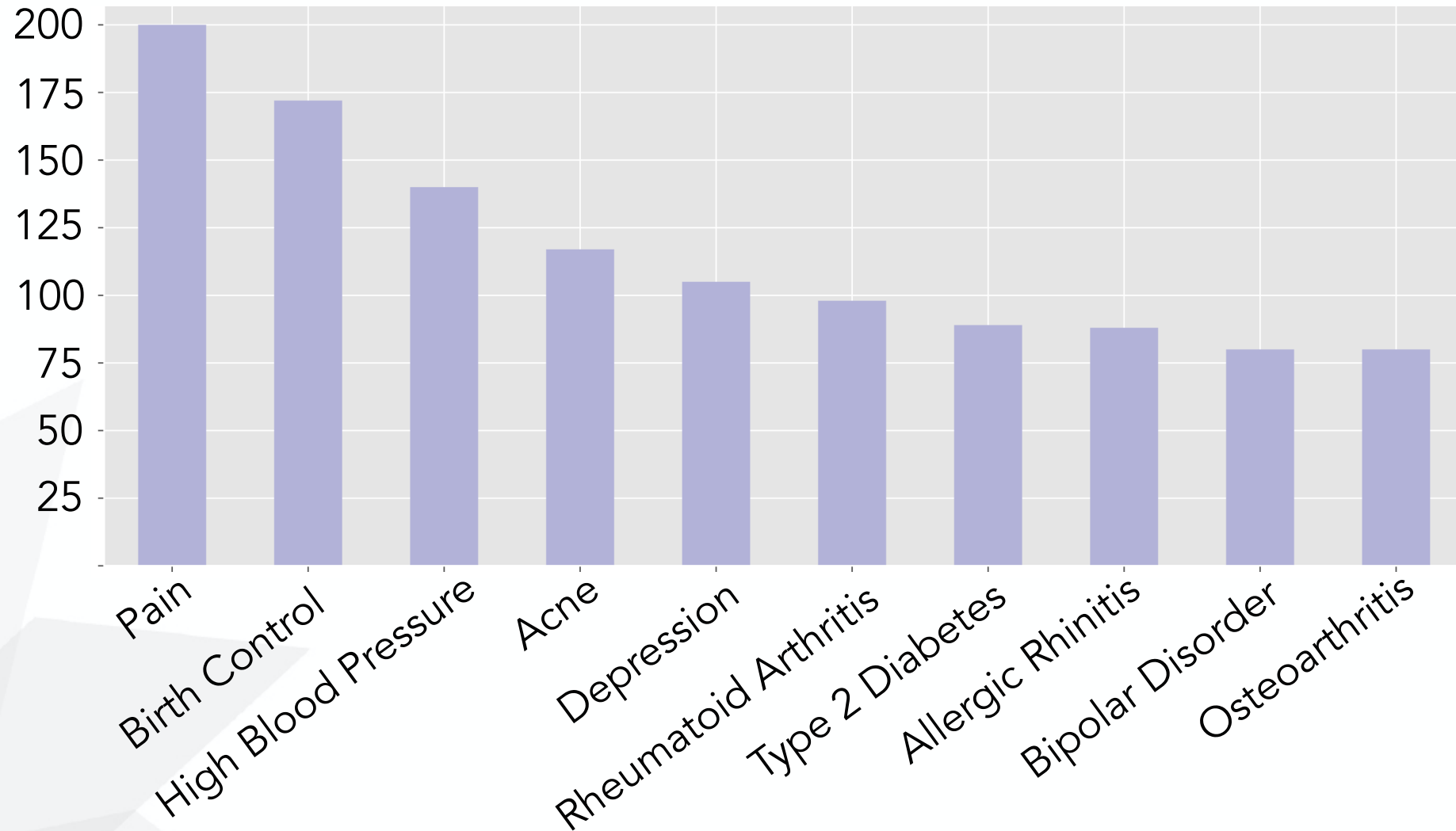


JiEun Song



Problem  
in drug market?

# Top 10 Conditions with Most Number of Drugs



# What makes a drug review useful?

---

- Too many different drugs for a condition
- Patients often have very little information

➔ Help patients' decision based on their concerns

# Workflow

Data  
Collection



UCI machine  
learning  
repository

Preprocessing



Clean data  
Remove stop words  
Lemmatizing  
CountVectorizer

Topic  
Modeling



Latent Dirichlet  
Allocation

Visualization



Python  
Tableau

# Topic Modeling

## Side Effect

year old dry mouth  
lost lb no side  
**side effect**  
lost pound blood pressure  
started taking  
year ago  
weight loss

## Feeling

took pill  
feel like within hour  
**felt like**  
hour later  
unprotected sex  
**first time**  
**yeast infection**  
two day first dose

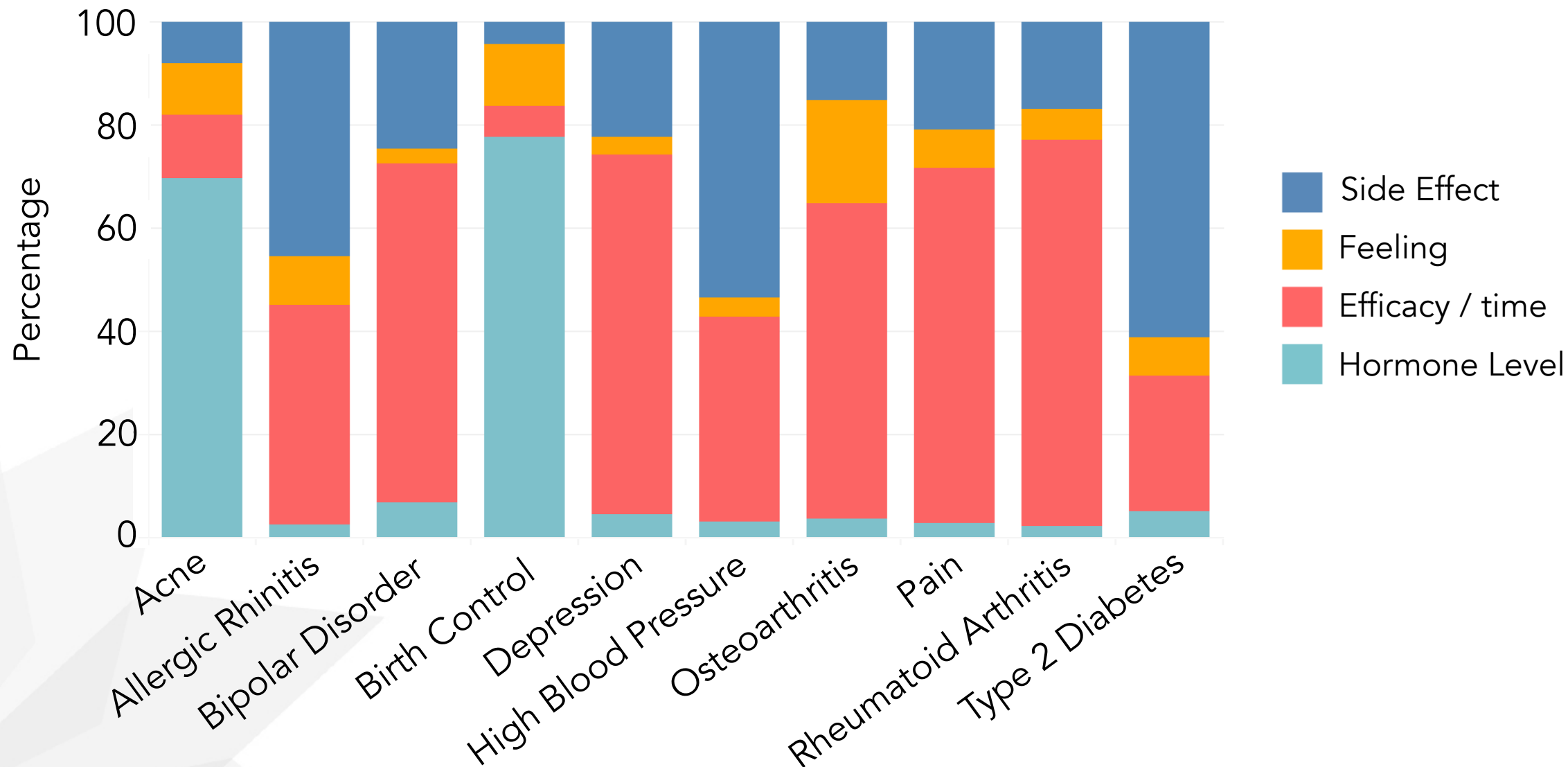
## Efficacy / Time

started taking  
**panic attack**  
no longer  
much better  
depression anxiety could not  
year old year ago  
back pain  
**feel like**

## Birth Control

gained weight  
no weight **weight gain**  
taking pill  
**birth control**  
first month  
mood swing  
sex drive  
side effect  
gained pound

# Topic Distribution of Top 10 Conditions



# Business Value and Future Work

---

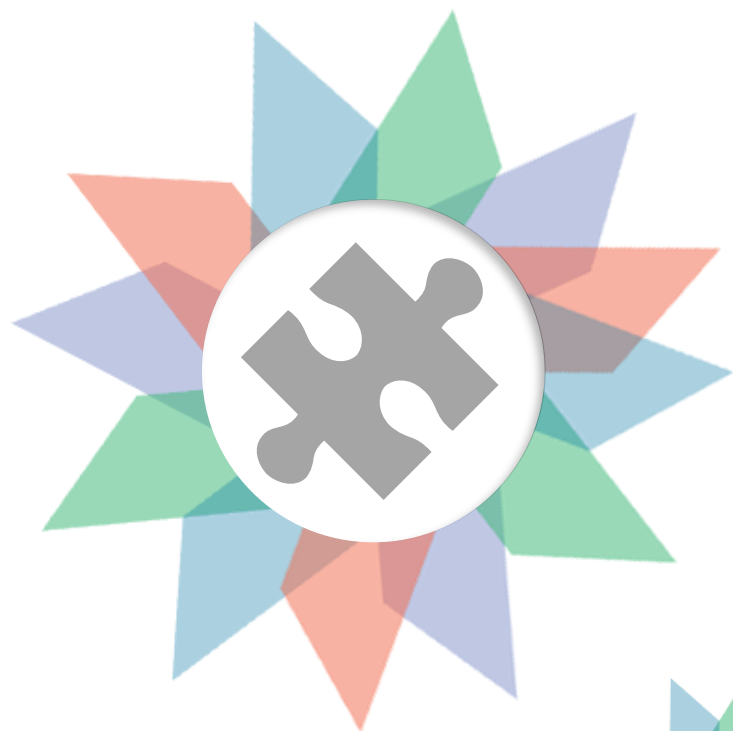
- Build a recommender system:
  - Help patients' decision making
  - Provide benchmark to drug providers
- Topic modeling for reviews across different conditions





**Thank you!**

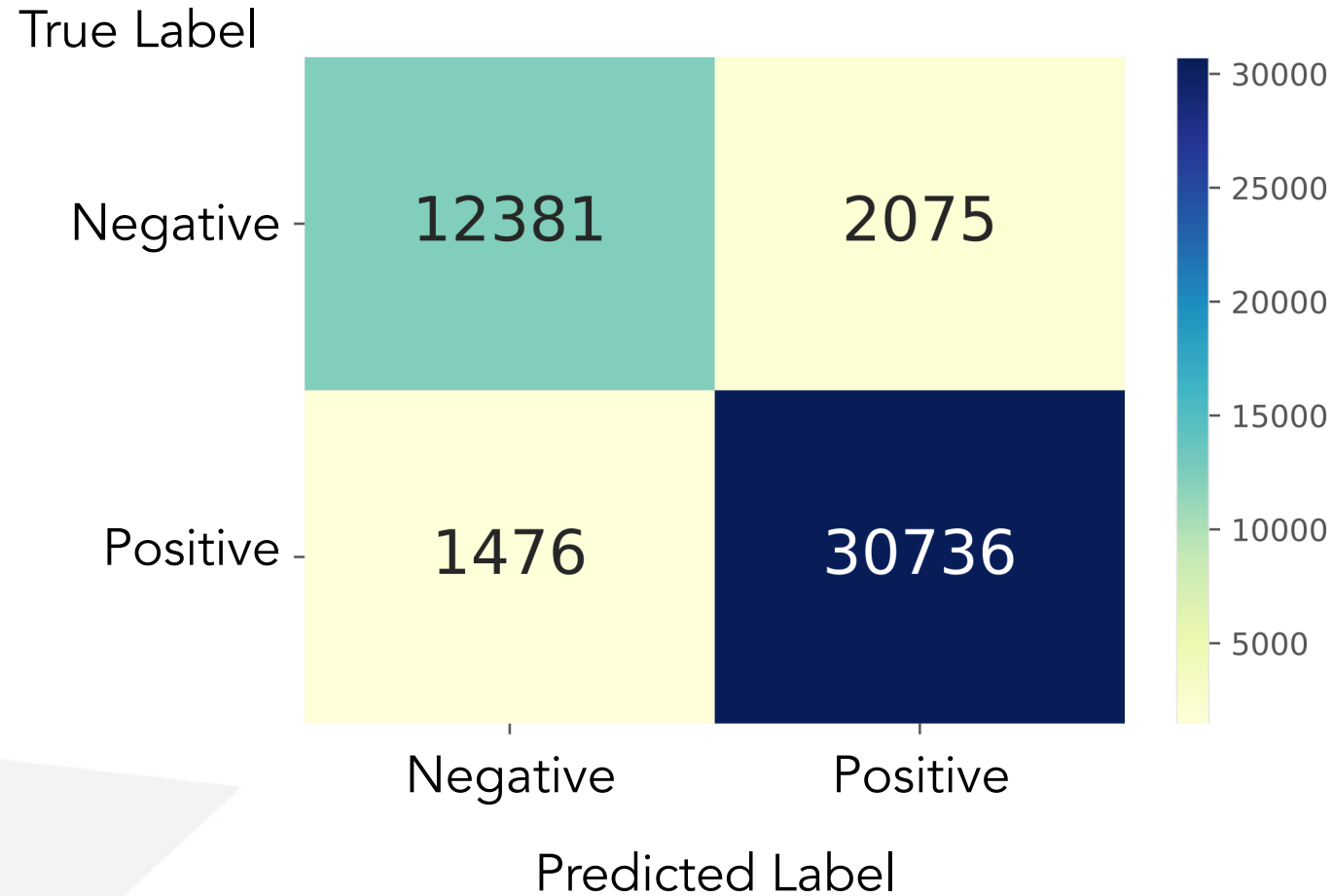
Question?



# Appendix



# Predicting Rating – Logistic Regression



Accuracy Score 0.92