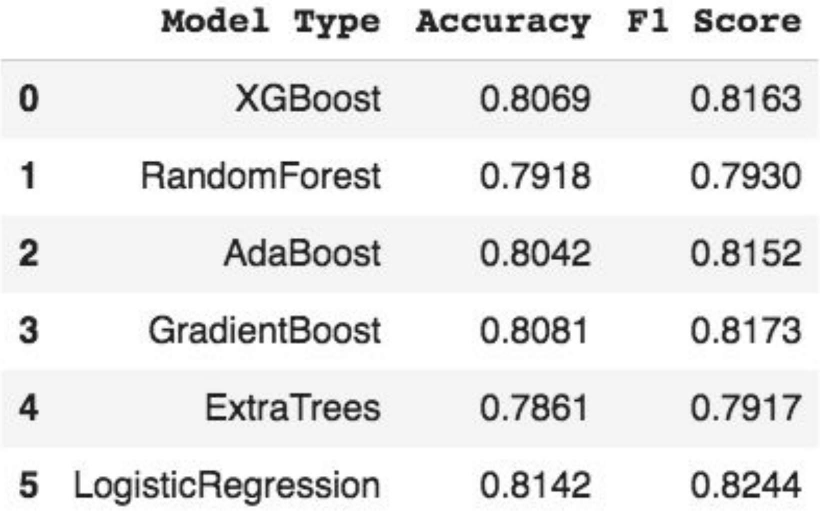
Social Media Determinants of Health

Marcus DeMaster, JingJing Rong, Johnny Yeo

Scoring Model Improvement

- Simple sequential neural net (~70% accuracy)
  - Trained on tweet text alone
  - No improvement with glove embeddings, LSTM, CNN nets
- NN output probabilities, profile features fed to ML Models.
- Ensembling: no improvement

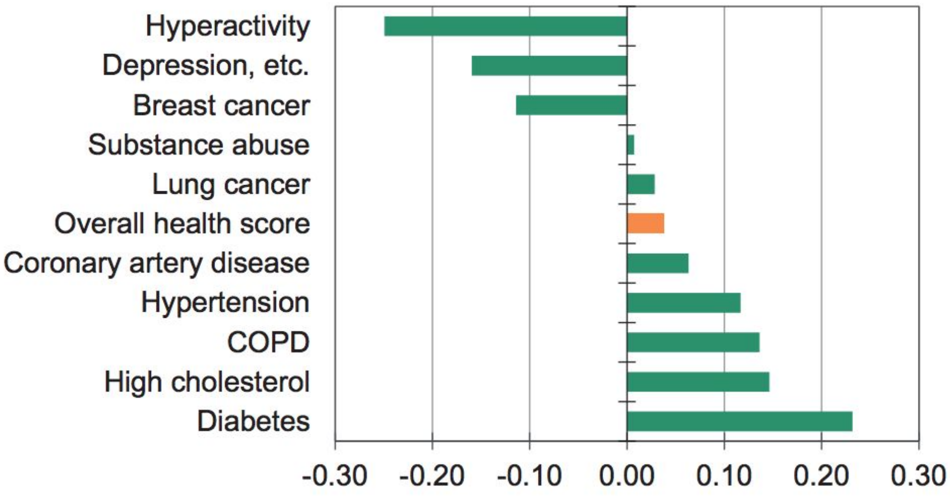


	Model Type	Accuracy	F1 Score
0	XGBoost	0.8069	0.8163
1	RandomForest	0.7918	0.7930
2	AdaBoost	0.8042	0.8152
3	GradientBoost	0.8081	0.8173
4	ExtraTrees	0.7861	0.7917
5	LogisticRegression	0.8142	0.8244

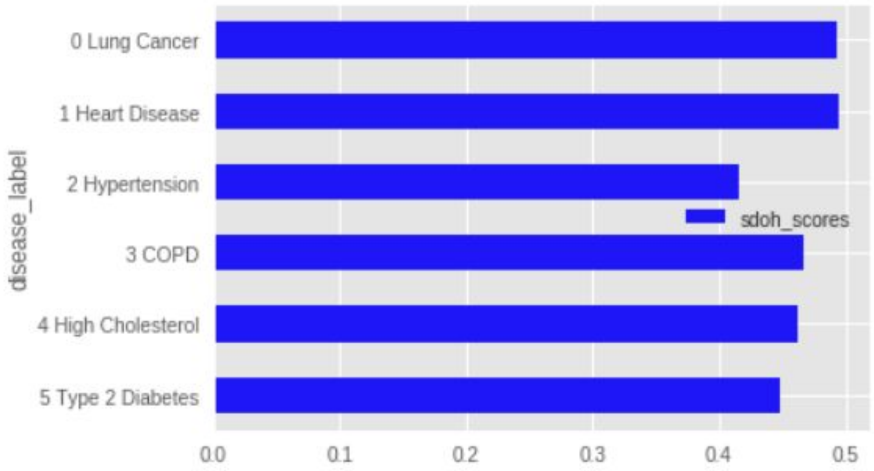
BCBS Trend vs. Mean Scores by Disease Group

Chart E2: Education Has Mixed Effects

Effect of % population with college degree on condition z-score

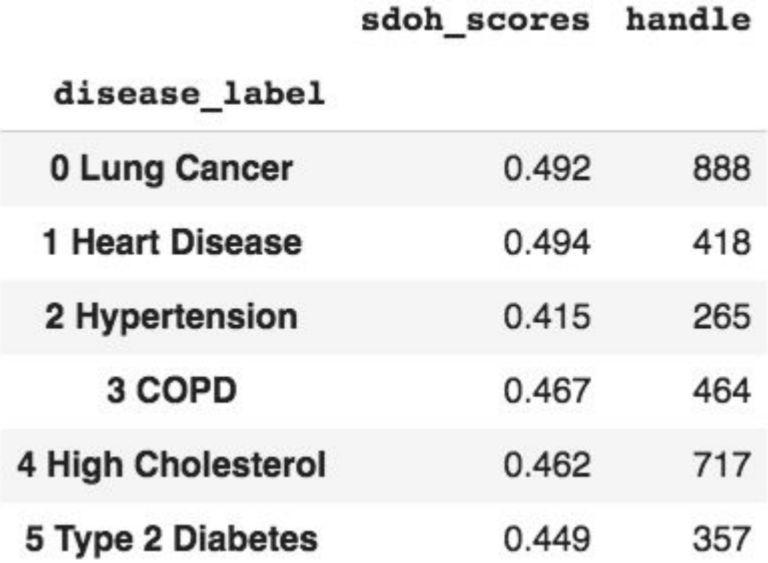


Sources: BCBS, Moody's Analytics



Disease Group Breakdown

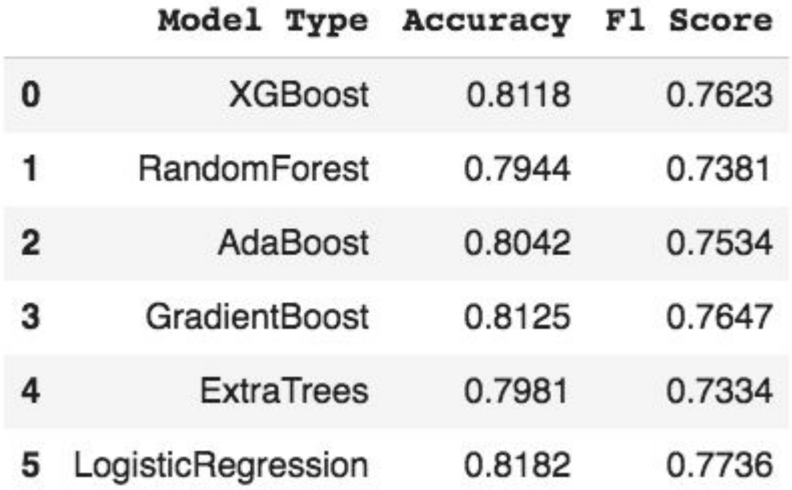
- Small Disease Group Size
- Group Size Varies
- More profiles to be collected for final dataset



	sdoh_scores	handle
disease_label		
0 Lung Cancer	0.492	888
1 Heart Disease	0.494	418
2 Hypertension	0.415	265
3 COPD	0.467	464
4 High Cholesterol	0.462	717
5 Type 2 Diabetes	0.449	357

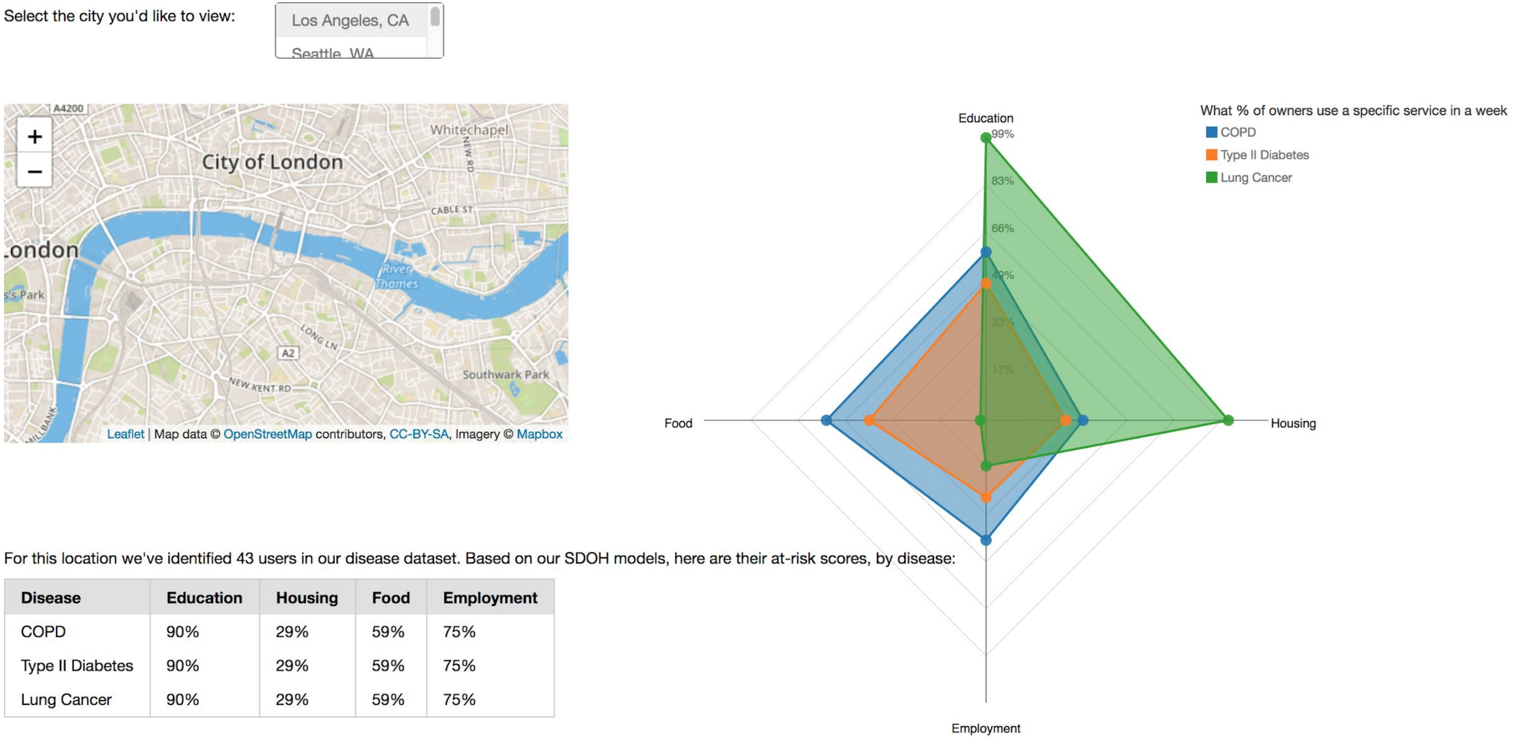
Employment Scoring Model

- Collected User Profiles
  - ‘I don’t have a job’ Profiles (1507)
  - ‘I have a job’ Profiles (2088)
- Manual Review
  - Removed ~14% of irrelevant users
  - ‘I don’t have a job’ Profiles (1286)
  - ‘I have a job’ Profiles (1804)
- Trained Model Performance
  - Neural Net + ML Classifier



	Model Type	Accuracy	F1 Score
0	XGBoost	0.8118	0.7623
1	RandomForest	0.7944	0.7381
2	AdaBoost	0.8042	0.7534
3	GradientBoost	0.8125	0.7647
4	ExtraTrees	0.7981	0.7334
5	LogisticRegression	0.8182	0.7736

Web Tool

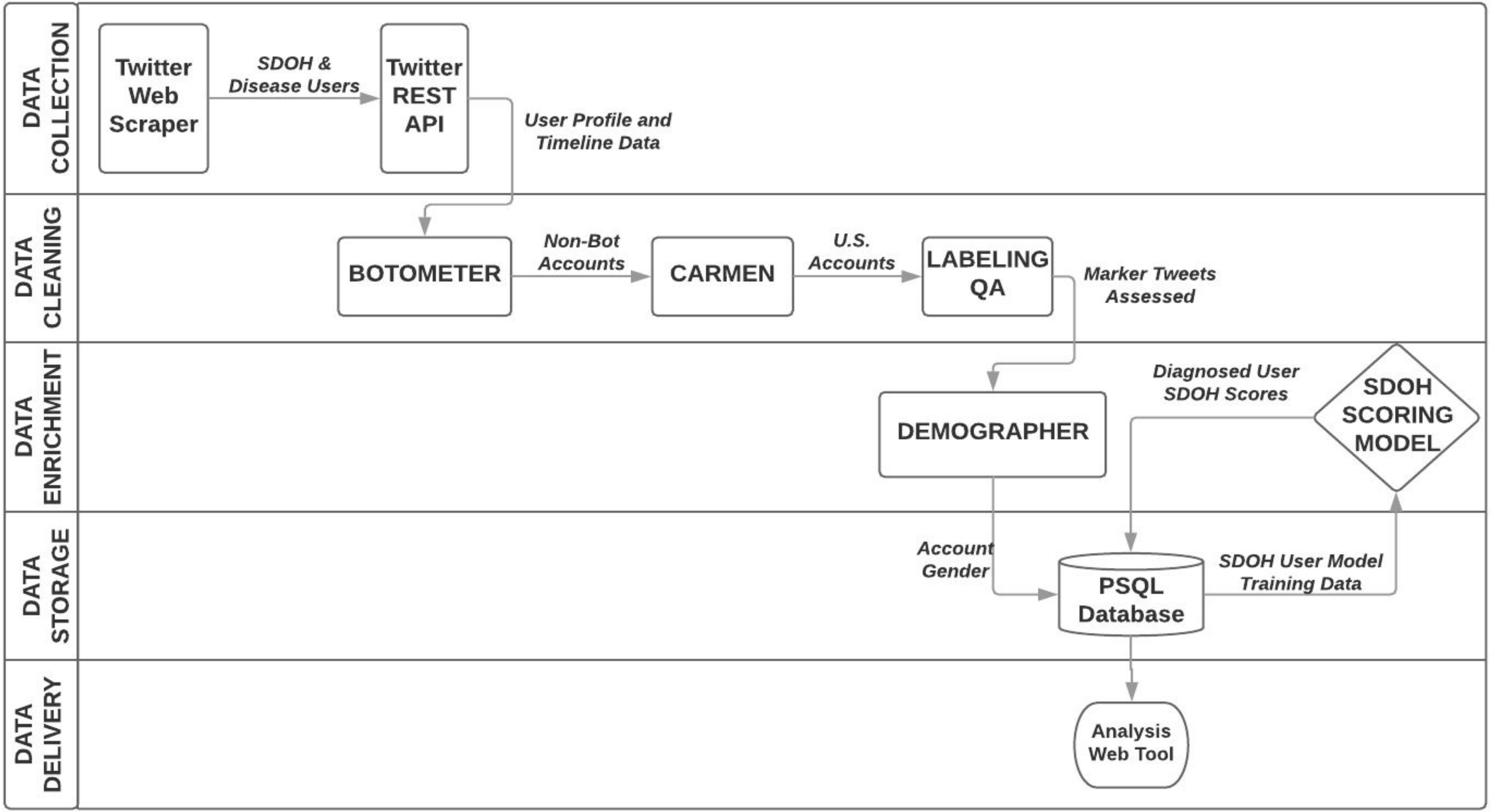


Select the city you'd like to view:

Leaflet | Map data © OpenStreetMap contributors, CC-BY-SA, Imagery © Mapbox

For this location we've identified 43 users in our disease dataset. Based on our SDOH models, here are their at-risk scores, by disease:

Disease	Education	Housing	Food	Employment
COPD	90%	29%	59%	75%
Type II Diabetes	90%	29%	59%	75%
Lung Cancer	90%	29%	59%	75%


DATA COLLECTION
DATA CLEANING
DATA ENRICHMENT
DATA STORAGE
DATA DELIVERY
Twitter Web Scraper
SDOH & Disease Users
Twitter REST API
User Profile and Timeline Data
BOTOMETER
Non-Bot Accounts
CARMEN
U.S. Accounts
LABELING QA
Marker Tweets Assessed
DEMOGRAPHER
Diagnosed User SDOH Scores
SDOH SCORING MODEL
Account Gender
PSQL Database
SDOH User Model Training Data
Analysis Web Tool