



# Data-driven Analysis of Equity and Fairness in Public Transit

Yihan Shao

PI: Dr. Ayan Mukhopadhyay  
Dr. Abhishek Dubey

# Introduction & Fairness in ML

## Task

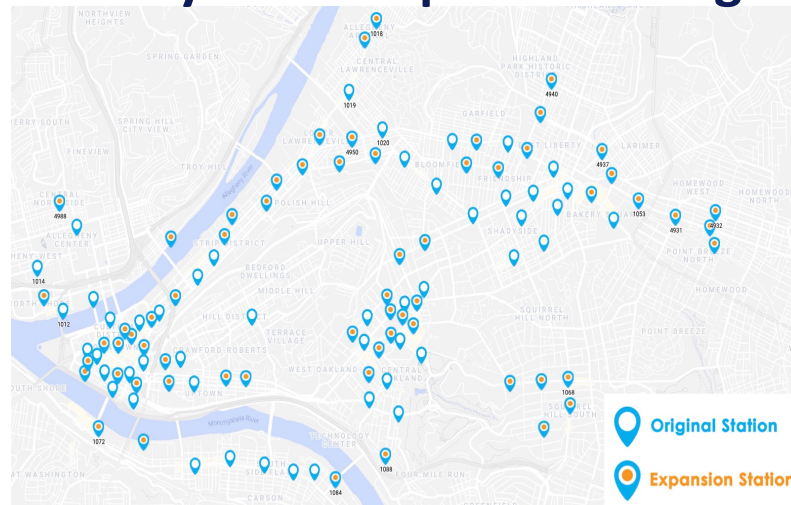
- predict the bike demand for the Healthy Ride with historical outflow data while ensuring relative fairness between the high-income and low-income groups.

## Fairness

Motivation: AI is objective only in the sense of learning what human teaches. The data provided by human can be highly-biased.

Ex. XING, a job platform similar to Linked-in, was found to rank less qualified male candidates higher than more qualified female candidates.

## Healthy Ride Map in Pittsburgh

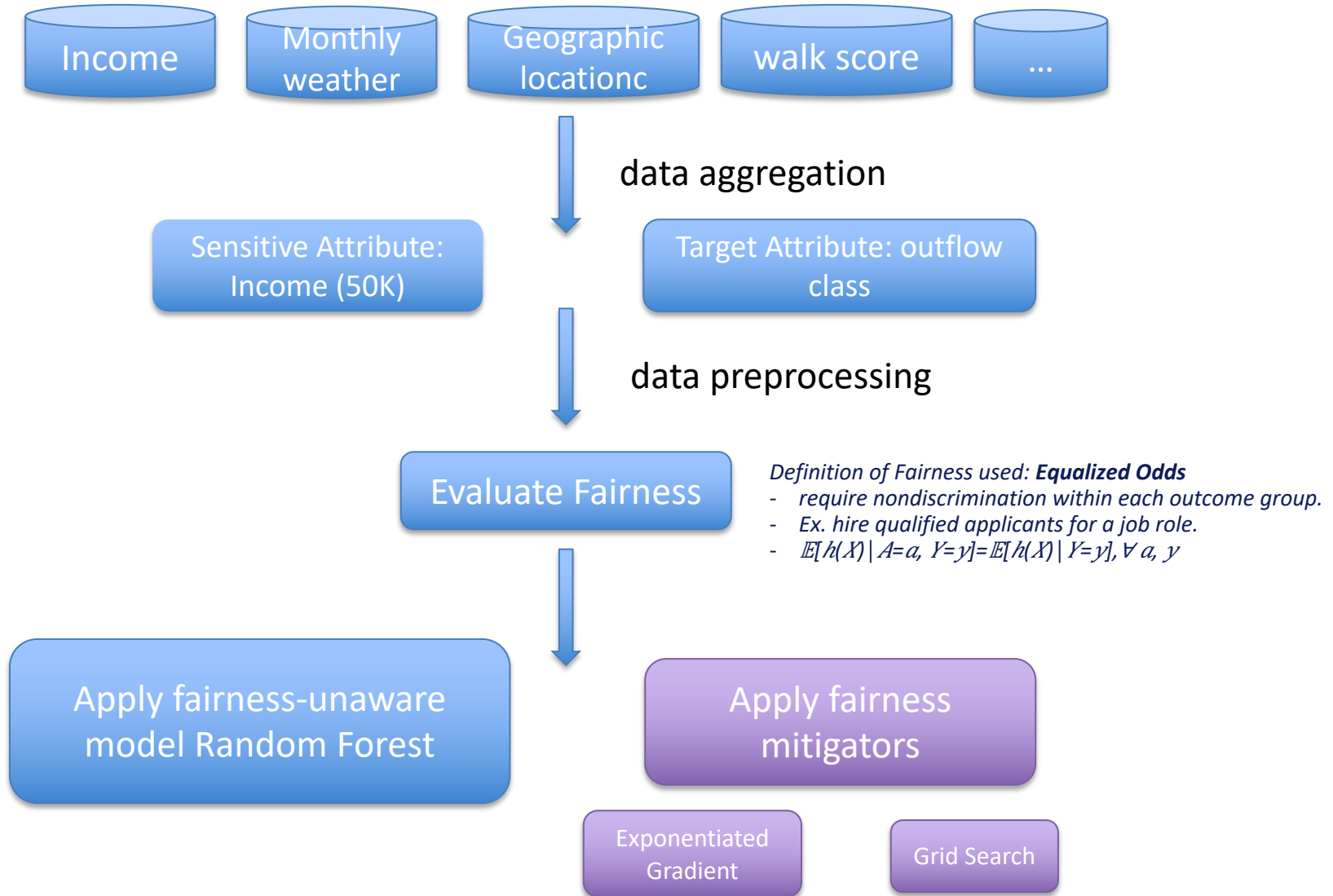


Search query	Work experience	Education experience	Profile views	Candidate	Xing ranking
Brand Strategist	146	57	12992	male	1
Brand Strategist	327	0	4715	female	2
Brand Strategist	502	74	6978	male	3
Brand Strategist	444	56	1504	female	4
Brand Strategist	139	25	63	male	5
Brand Strategist	110	65	3479	female	6
Brand Strategist	12	73	846	male	7
Brand Strategist	99	41	3019	male	8
Brand Strategist	42	51	1359	female	9
Brand Strategist	220	102	17186	female	10

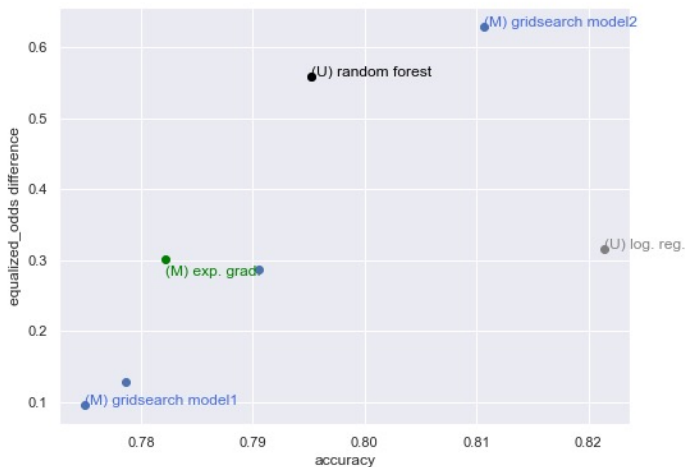
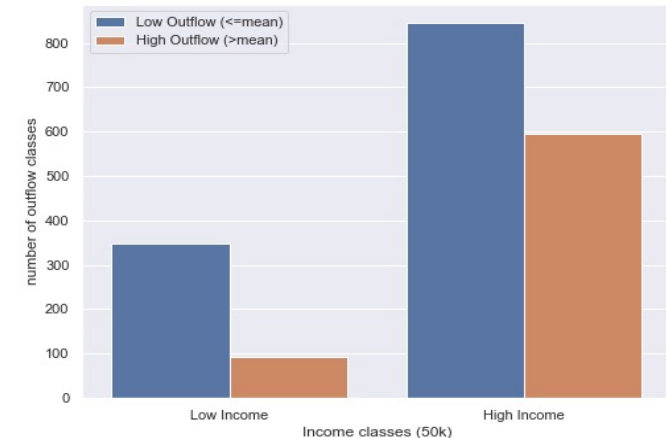
TABLE II: Top k results on [www.xing.com](http://www.xing.com) (Jan 2017) for the job search query “Brand Strategist”.

Figure: The bias in the query for Brand Strategist from XING(from [Lahoti et al. 2018](#)).

# Workflow



# Model results



- **Equalized odds difference:** implemented as the greater value of the true positive rate difference and false positive rate difference.
- **Grid search:** implemented as a simplified version of the exponentiated gradient reduction and works by generating a sequence of relabelings and reweightings, and trains a predictor for each.
- **Ideal case:** a predictor at (1,0) - perfectly accurate and without any unfairness under equalized odds (with respect to the sensitive feature “income class”).
- The range of the axes - the disparity axis covers more values than the accuracy. Can reduce disparity substantially for a small loss in accuracy.
- Pick the model which represented the best trade-off between accuracy and disparity given the relevant business constraints.

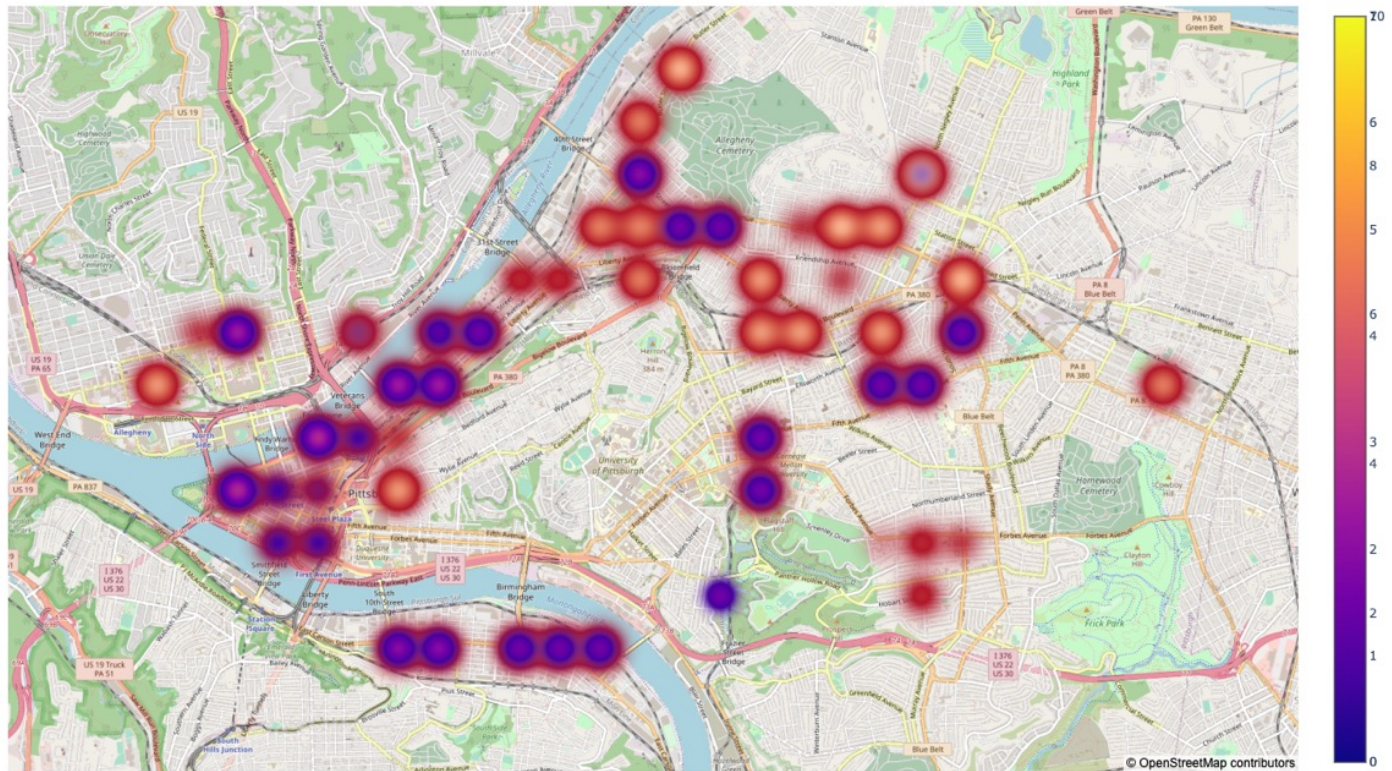
# Model Visualization

Geological visualization of predicted bike demand for Pittsburgh  
: income and Unmitigated Prediction of # of high demand groups

Income & Unmitigated prediction ▼

**Red dot:**  
distribution of  
high-income  
classes

**Purple dot:**  
distribution of  
high-outflow  
classes



**Unmitigated Prediction**



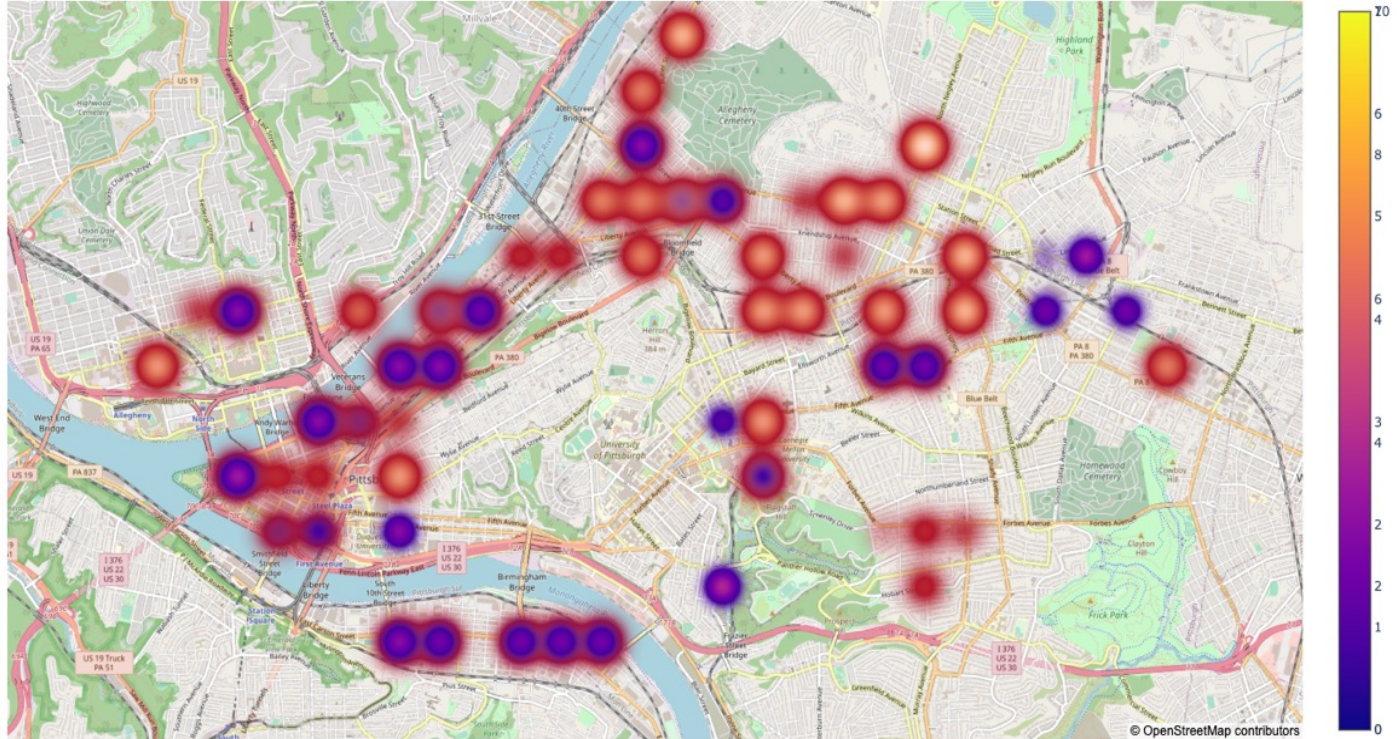
# Model Visualization

Geological visualization of predicted bike demand for Pittsburgh  
: income and Mitigated Prediction with Grid Search Model 1 (low disparity) of # of high demand groups

Income & Mitigator gridsearch  
model 1 (low disparity) ▼

**Red dot:**  
distribution of  
high-income  
classes

**Purple dot:**  
distribution of  
high-outflow  
classes



**Mitigated Prediction**

# Conclusion

- Lessons:
  - It's not only being “different” is an effective result, being the “same” is also a result.
  - Pay more attention when preprocessing data.
  - Learn new ideas and apply them quickly
- Challenges
  - Bugs & debugging
- Things go well:
  - I will be continuing research in the Fall semester.