EUR/USD - 1,35379 - 00:00:00 14 giu (EEST)
EUR/USD (Bid), Ticks, # 300 / 300
23:35
23:40
23:46
Gold, spot - 1,276,820 - 23:00:00 13 giu (CEST)
Gold, spot (Bid), 1 minute, # 159 / 300, Logarithmic, Heikin Ashi
Settings
Quote List

NXTGEN SUMMER ACADEMY

Data Science for Good

July 27, 2021

Presentation Outline

Today's Discussion

What is data science for good?

How do we get involved?

Open discussion - get to know us!

LEARNING OBJECTIVES

✓ **Recognize**
societal and community-driven issues that can benefit from data-driven support

✓ **Demonstrate**
knowledge of how data science is applied to real-world problems

✓ **Gain**
practical experience with problem analysis and decision making

✓ **Understand**
your communities and your potential roles as data for good advocates

Why Data for Good?

A Brief Background

Volunteering your time for the good of others can take many forms

Just like Teach for America or Doctors without Borders, data scientists can donate their time and skills in meaningful ways

Non-profit organizations and humanitarian causes, for example, can benefit from answering data-driven questions

Data Science for Good

Why it Matters

Community organizations can use data to better share their resources, connect with the people they serve, and improve their goals

Causes can use data to define key problems, determine how to answer important questions, and understand how their impact changes over time

Researchers can use data to test hypotheses, predict outcomes, and uncover relationships between the things they study

CONSIDER...

How to...

Determine the best routes and parking places for mobile food pantries in low-income areas, to reach the most people in need?

How to...

Predict whether a given imported shipment that contains wildlife or wildlife products is illicit or not, to assist local and global government agencies stop wildlife trafficking?

How to...

Improve election polling and survey methodology for presidential elections using social media data?

How to...

Computationally identify racial stereotypes and biases in written text to mitigate health consequences for people of color in a digital world?

THE BEST THING ABOUT BEING A [DATA SCIENTIST] IS THAT YOU GET TO PLAY IN EVERYONE’S BACKYARD.

- *John Tukey*

Stephen Salerno

salernos@umich.edu

Who am I?

I am a PhD student in biostatistics at UM, studying patient survival and how measures of healthcare quality are publicly reported

What was my spark?

First research project developing methods / analyzing data on how tuberculosis can be detected in endemic areas

How do I get involved?

Volunteer with Statistics in the Community (STATCOM), a community outreach program for data science consulting

Ani Madurkar

amadurka@umich.edu

Who am I?

Data Scientist/Engineer at Jackson and Applied Data Science Masters graduate from UM. I build robust, automated systems and tell stories of insights for business leaders to take action on.

What was my spark?

Conducting data analysis on brain scans for psychiatric patients in a neuroscience research lab in undergrad at Wayne State.

How do I get involved?

Partner with researchers to work on data for good projects and write stories on Medium

Katie Nicholson

knich@umich.edu

Who am I?

I am an undergraduate student in Data Science Engineering at UM, hoping to work in industry after graduation

What was my spark?

Autonomous drones class freshman year introduced me to data and how it can be used to make driving safer

How do I get involved?

Research in election polling and survey methodology to make presidential election polls more accurate and accessible

Gauri Kambhatla

gkambhat@umich.edu

Who am I?

I am a PhD student in artificial intelligence (NLP) at UT, and just finished my undergraduate and Master's at UM.

What was my spark?

Cognitive Science undergrad course, learning about the human mind and the complexities of how we learn language

How do I get involved?

Research in the AI lab at UM, specifically the Language and Information Technologies (LIT) lab

Data Science for Good

Example Projects

Food for Thought, Toledo

Global Wildlife Trafficking

Improving Election Polling

Surfacing Racial Stereotypes

Food for Thought, Toledo

Optimizing Mobile Food Pantry Locations

Background

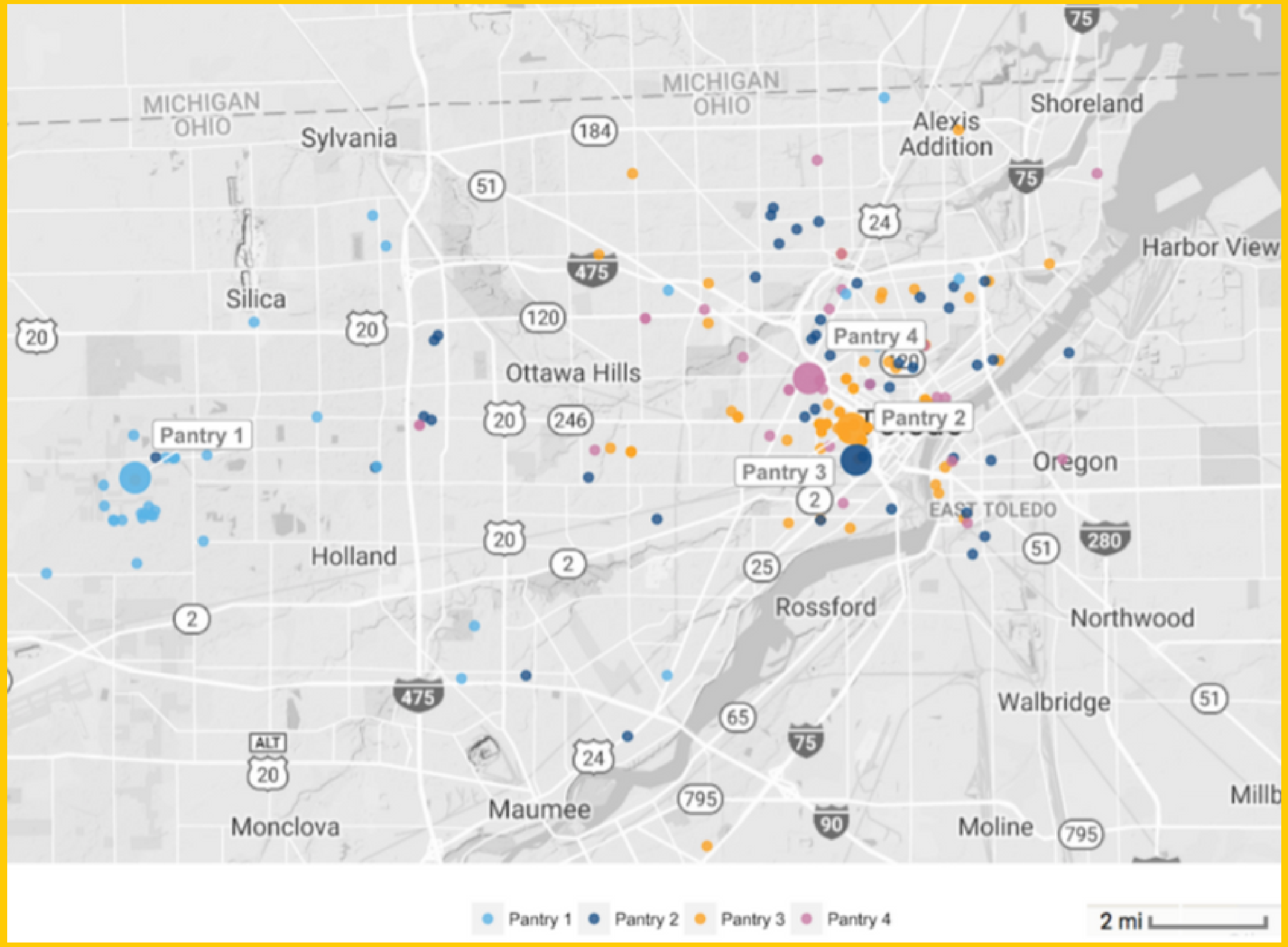
FFT is a non-profit that serves 400+ families experiencing food insecurity in Toledo, Ohio each month through a mobile pantry service

Goal

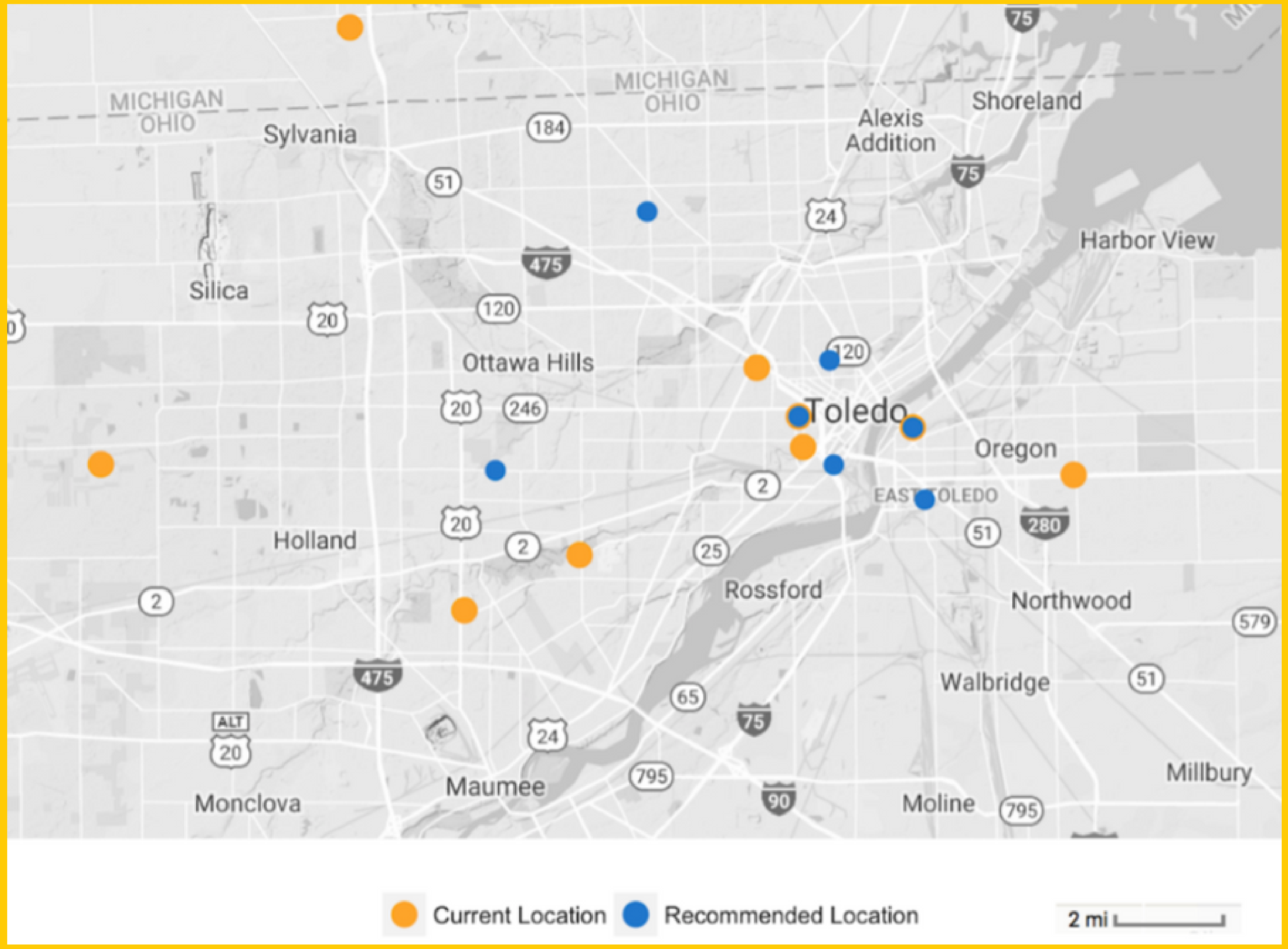
Find the optimal pantry locations and order they should be visited to best allocate resources to households across the city within a given month

Result

Food for Thought, Toledo is now scheduling routes for their mobile pantries based on our recommendations and is now reaching more people than ever before



Previous Pantry Locations



Current and Optimal Locations

Analyzing Global Wildlife Trafficking

Using Graph-Based Methods and Machine Learning to Assist Researchers & Officers Detect Patterns in Data

Background

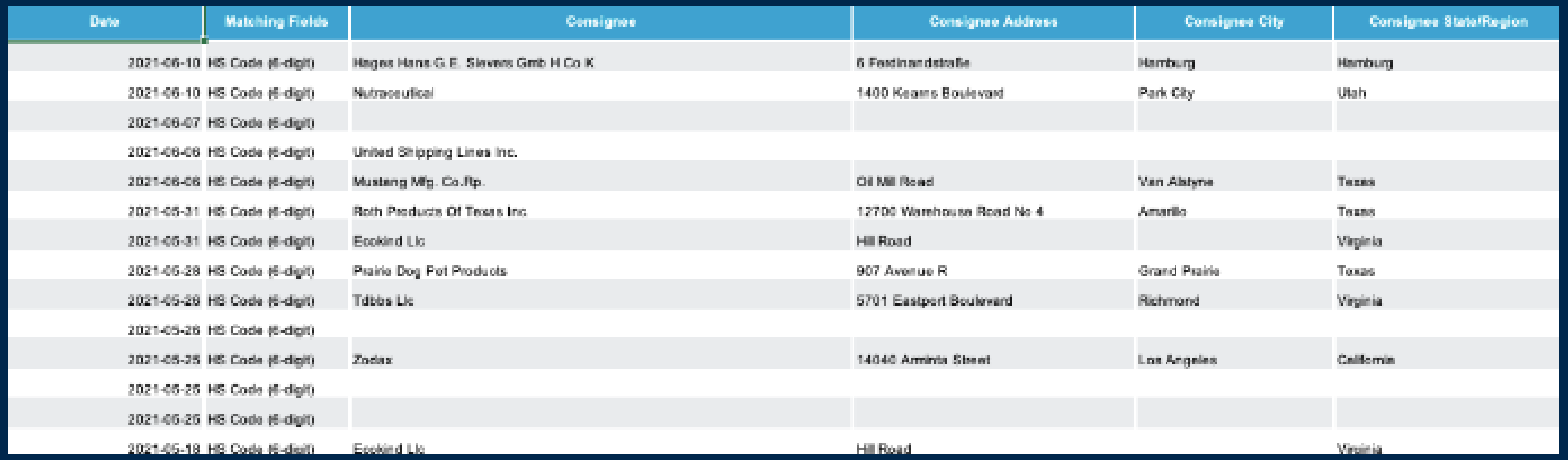
Wildlife Trafficking is a multi-million dollar problem that undermines security problems across nations. Even so, researchers and officers don’t have a systematic way of targeting & preventing illicit activity

Goal

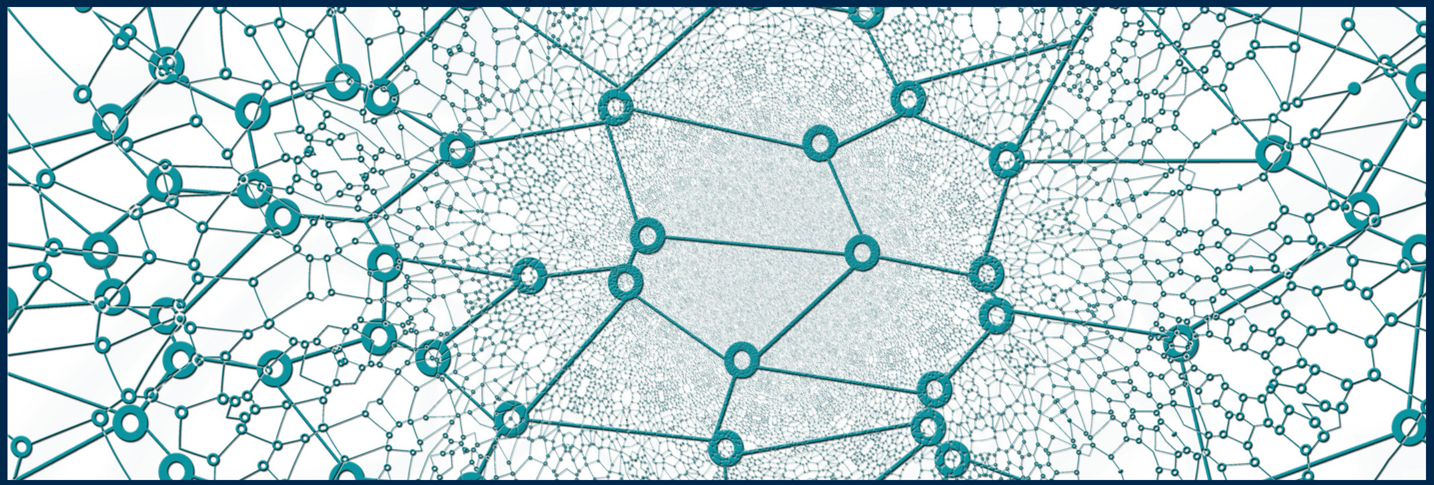
Create a dashboard that automatically reads in new data of wildlife/wildlife product shipments and provides insight into the ‘riskiest’ shipments that need to be checked

Result

Currently still in progress!



Date	Matching Fields	Consignee	Consignee Address	Consignee City	Consignee State/Region
2021-06-10	HS Code (6-digit)	Hagas Hans G.E. Sievers Gmb H Co K	6 Ferdinandstraße	Hamburg	Hamburg
2021-06-10	HS Code (6-digit)	Nutraceutical	1400 Kearns Boulevard	Park City	Utah
2021-06-07	HS Code (6-digit)				
2021-06-08	HS Code (6-digit)	United Shipping Lines Inc.			
2021-06-06	HS Code (6-digit)	Mustang Mfg. Co.ltp.	Oil Mill Road	Van Alstyne	Texas
2021-05-31	HS Code (6-digit)	Roth Products Of Texas Inc.	12700 Warehouse Road No 4	Amarillo	Texas
2021-05-31	HS Code (6-digit)	Ecokind Llc	Hill Road		Virginia
2021-05-28	HS Code (6-digit)	Prairie Dog Pet Products	907 Avenue R	Grand Prairie	Texas
2021-05-28	HS Code (6-digit)	Tdbbs Llc	5701 Eastport Boulevard	Richmond	Virginia
2021-05-28	HS Code (6-digit)				
2021-05-25	HS Code (6-digit)	Zodax	14040 Arminta Street	Los Angeles	California
2021-05-25	HS Code (6-digit)				
2021-05-25	HS Code (6-digit)				
2021-05-18	HS Code (6-digit)	Ecokind Llc	Hill Road		Virginia



Media Mood

Improving Election Polling

Background

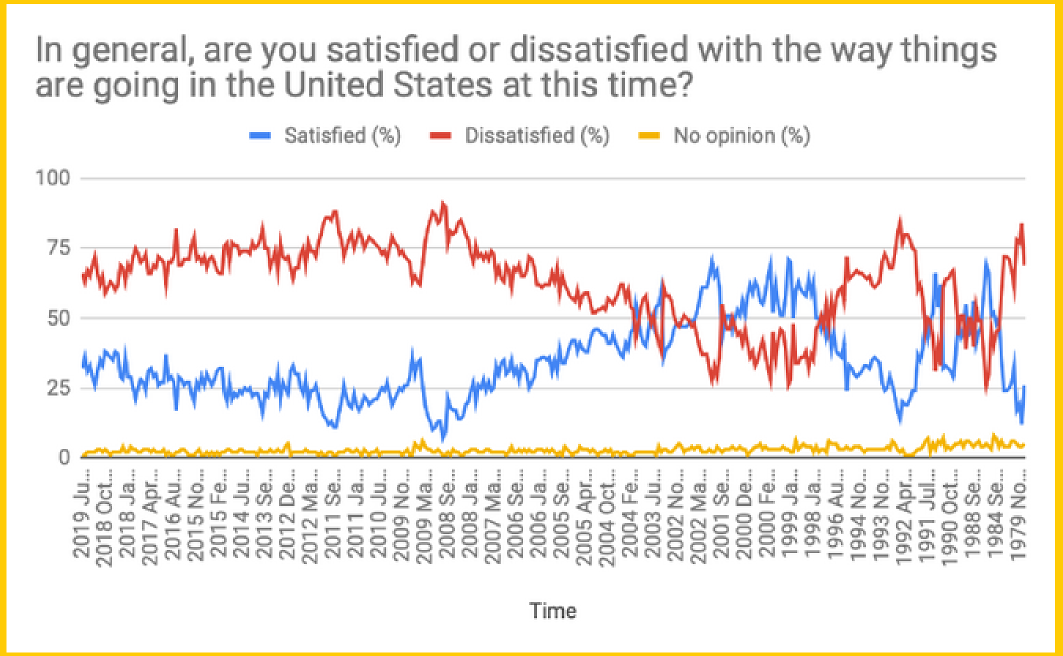
Public opinion polls are collected through phone surveys, which are not always accurate

Goal

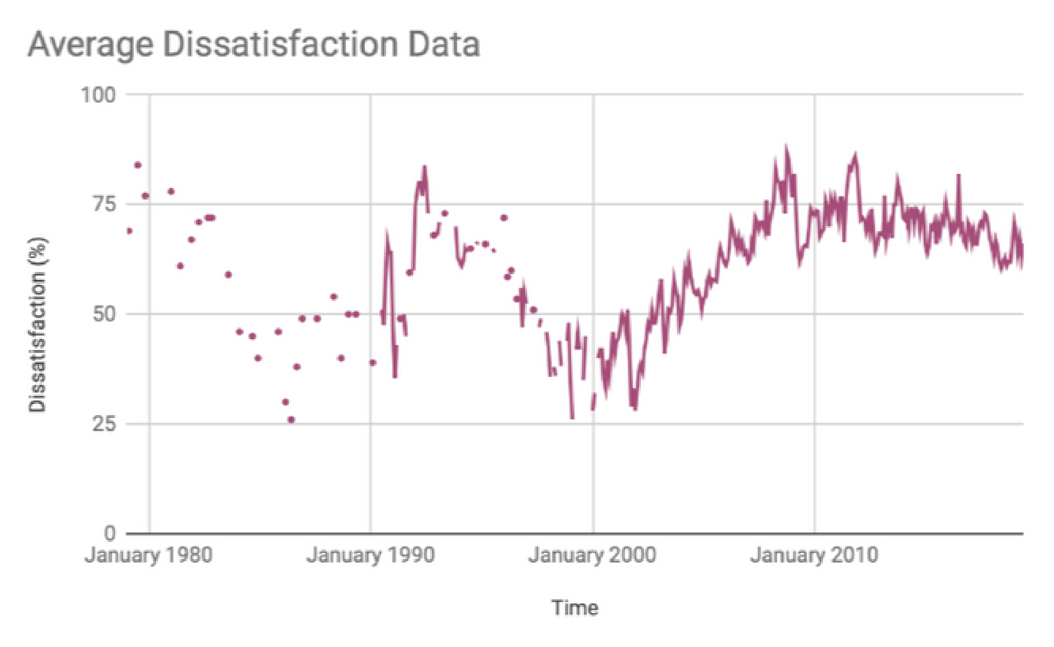
This project models previous data from past phone surveys and attempts to use other forms of media to replace phone surveys

Result

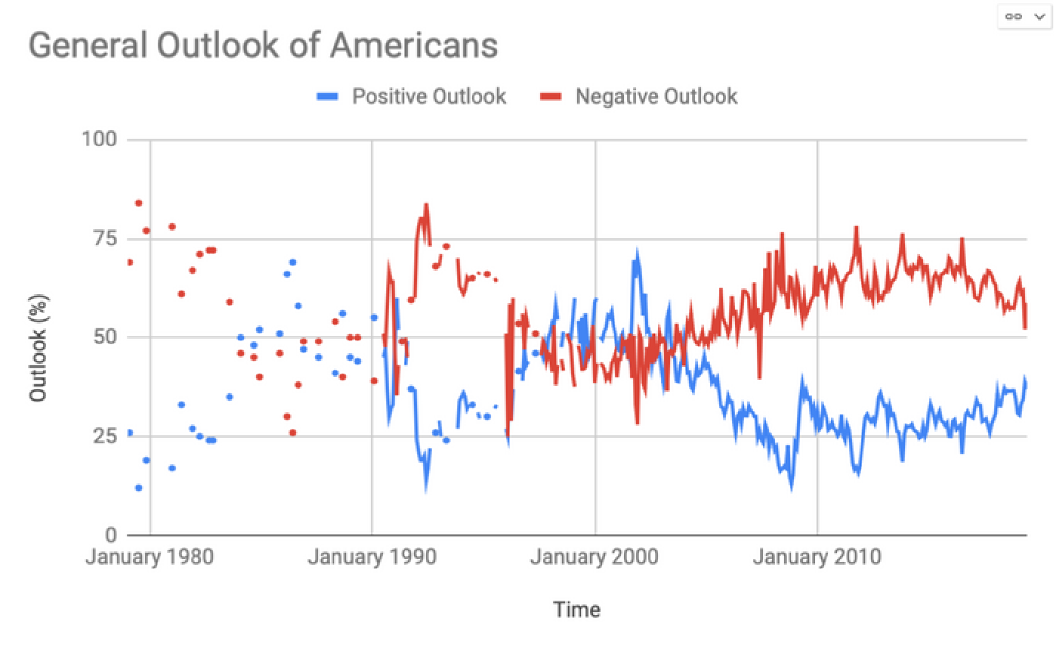
There are statistical trends in newspaper and social media data that indicate general mood indicators, which can be used to predict polls



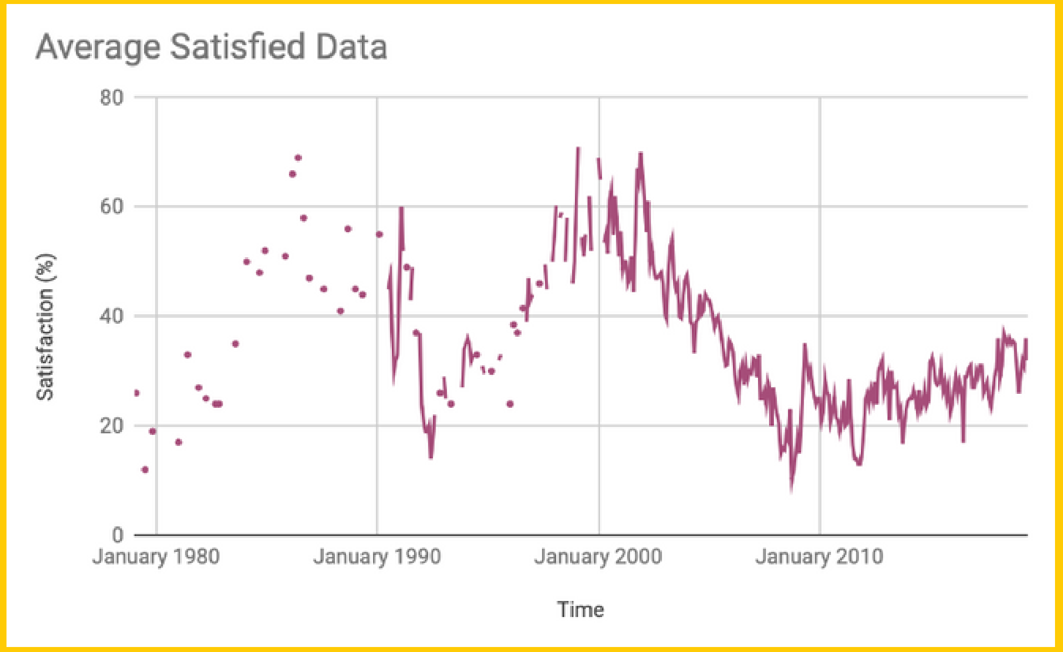
In general, are you satisfied or dissatisfied with the way things are going in the United States at this time?



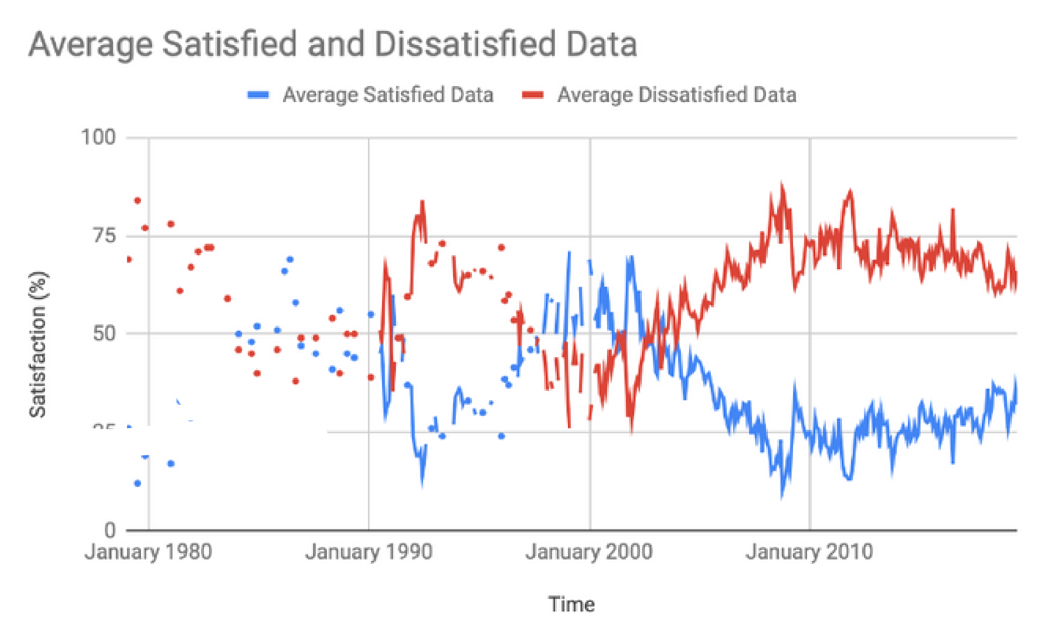
Average Dissatisfaction Data



General Outlook of Americans



Average Satisfied Data



Average Satisfied and Dissatisfied Data

Data Visualization

Surfacing Racial Stereotypes

How can we computationally identify racial stereotypes and biases?

Background

People of color deal with the consequences of implicit racial bias all the time, much of which is now online due to the vast amounts of digital interactions

Goal

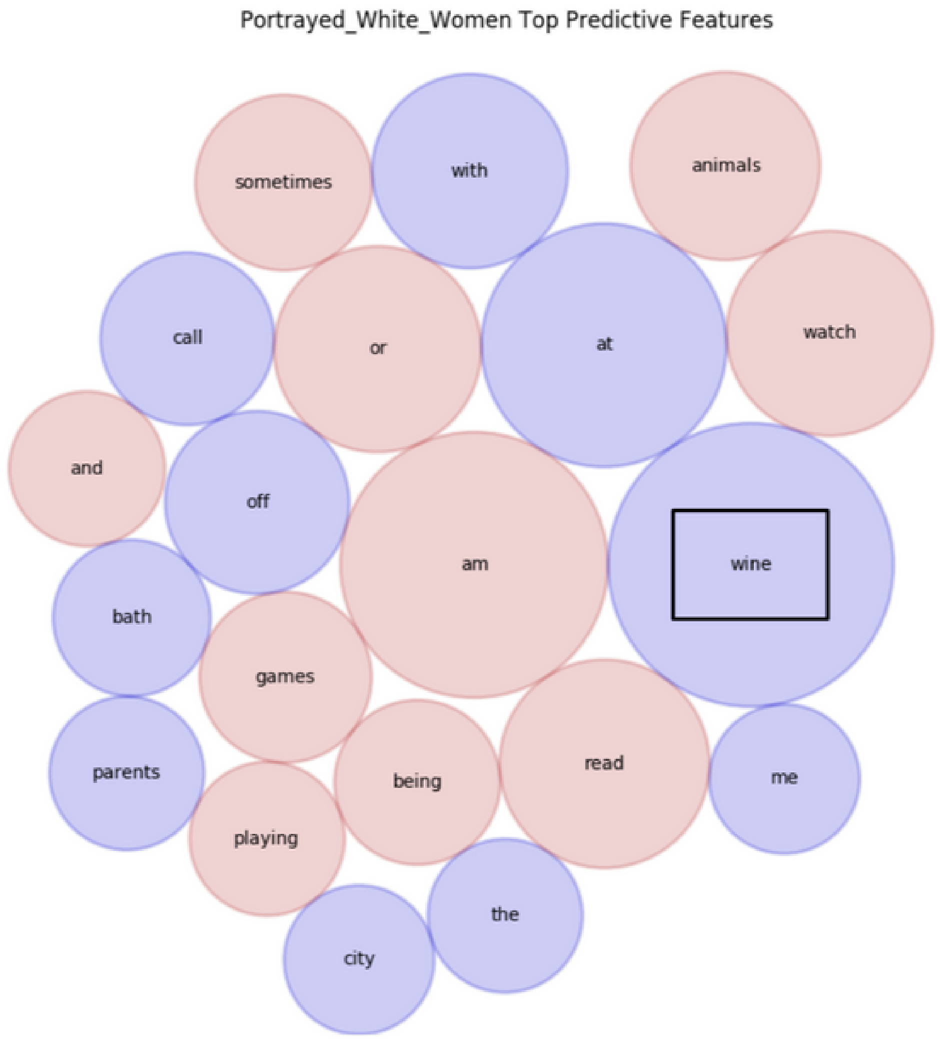
Long term:
Computationally identify racial stereotypes in text

This project:
explore how we can do the above through racial identity portrayal [Black/White women/men]

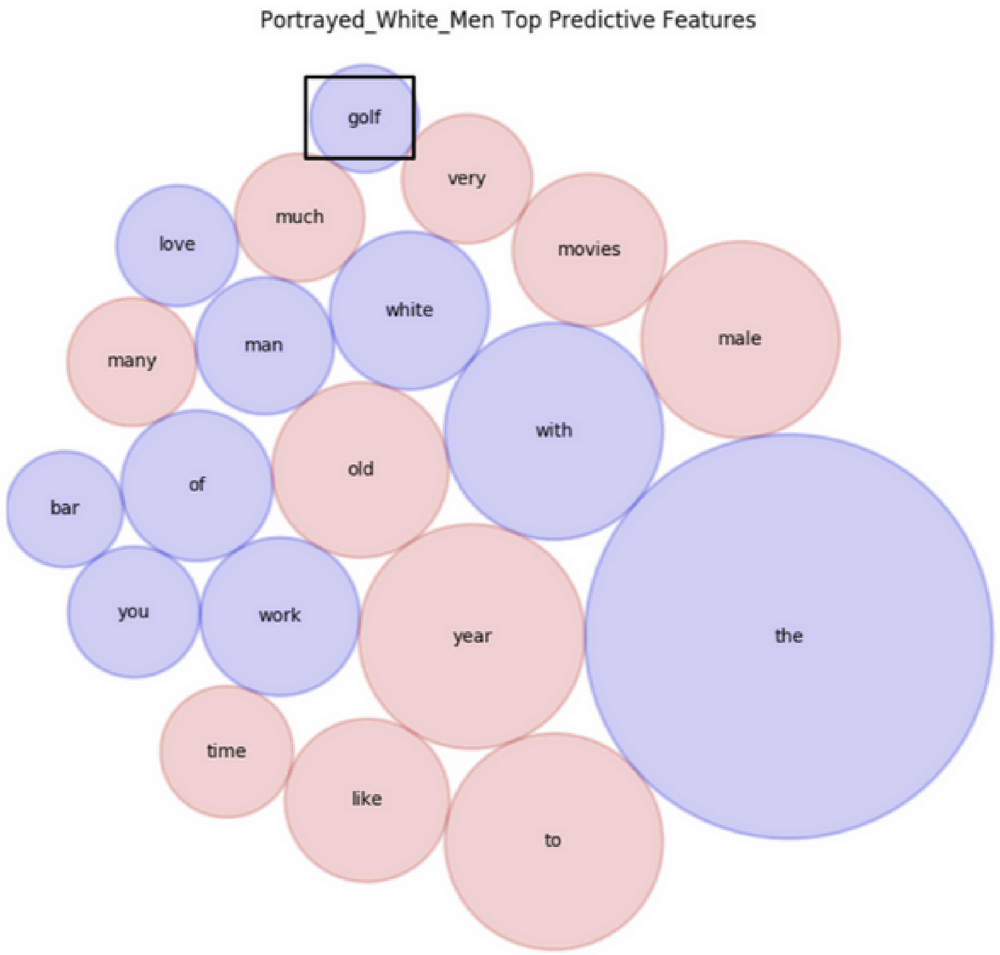
Result

Easier to classify portrayed Black identities than portrayed White identities, predictive linguistic features (words) reveal stereotypes / generalizations!

⋆work in progress!



Portrayed_White_Women Top Predictive Features



Portrayed_White_Men Top Predictive Features

Time For

Open Discussion

Get to know us!

NxtGen Summer Academy • July 2021

OPEN DISCUSSION

What

has been one of the most rewarding project you have worked on?

How

can you get involved with data science for good intitatives as a student?

Do

you have any advice for new students in data science or related majors?

Can

data science for good be a part of your life outside of school?

MICHIGAN INSTITUTE FOR DATA SCIENCE

Twitter

twitter.com/um_midas

E-Mail

midas-contact@umich.edu

Website

midas.umich.edu

Reach Us

For questions or more info:

Stephen Salerno

salernos@umich.edu

Ani Madurkar

amadurka@umich.edu

Katie Nicholson

knich@umich.edu

Gauri Kambhatla

gkambhat@umich.edu

THANK YOU!

Additional Slides

More Project Information

Food for Thought, Toledo

Global Wildlife Trafficking

Improving Election Polling

Surfacing Racial Stereotypes

Food for Thought

Background

FFT is a non-profit that serves 400+ families experiencing food insecurity in Toledo, Ohio each month through a mobile pantry

Goal

Having collected data on the performance of 35 different locations for several years, they sought to identify optimal pantry locations

The Data

When receiving food, families were asked to report an address and other demographic data such as how many people in the household

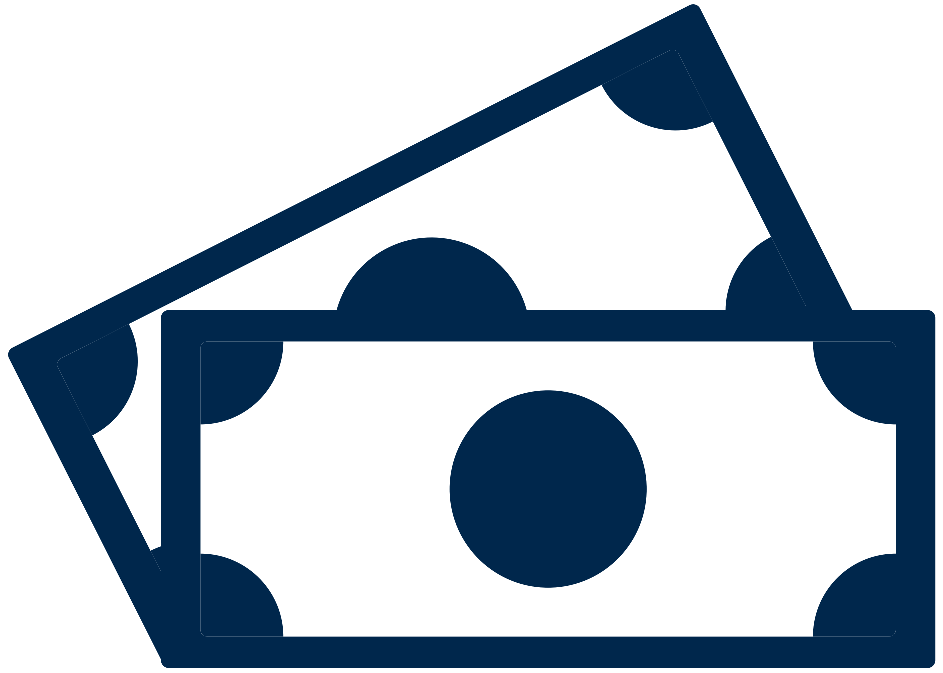
THE DATA

Demographics

Families were asked to report an address, and other demographic data such as how many people in their household

Health Outcomes

Several health characteristics and food preferences were also surveyed to better gauge the needs of those being served



SES

Neighborhood socioeconomic data were also collected using community surveys and governmental/census resources

Understanding the Data

Clustering analysis used to group/rank areas with similar characteristics into need categories; based on regional poverty, unemployment, and health characteristics. Distances traveled to a pantry used to minimize total distance traveled, constrained to higher priority neighborhoods and other limiting factors

Optimizing Pantry Locations

Location modeling used to create network of demand nodes, optimized locations of accruing unmet demand with constraints related to the number of visits per month and the needs of the households covered by a candidate location

A Good Project

Reasons Why

Oversaw from conception to interpretation

Findings used to inform decisions

Leveraged many partnerships

Students directly impacted this population

Surfacing Racial Stereotypes

Additional Details

Prompts

- Please describe yourself.
- Please describe your typical evening on a workday, after a day at work or school.
- Imagine you are a [fake identity], the same age as you. Please write from the first-person perspective of a [fake identity]. Others will read what you wrote; your goal is to convince them it was written from the perspective of a [fake identity] without saying so explicitly. For example, do not write a sentence like "I am a [fake identity]" as this is an explicit statement.

EXAMPLES

Real Black Woman

I am a Black woman born and raised in Alabama. I have been married for 5 years to my husband and we have a 2 year old daughter and 1 year old son. I am a stay at home mom. I am also disabled. I have two autoimmune disorders. My dimpled smile is my favorite thing about me...

Fake White Woman

I'm a southern soccer mom who loves to live, laugh, love! I have a wonderful husband and a dear son and daughter. I'm a blond little thing with a big smile and bigger hugs. I am the life of the party after a few glasses of wine. I go to yoga every morning to keep in shape...

Real White Woman

My name is Rachell. I'm a 33 year old white female and I live in Missouri. I am the first time mother to a 4 month old baby boy named Jax. I am a recovering addict. I start college on March 22nd. I've came a long way in my recovery. I'm in drug court. I like to listen to Christian music...

Fake Black Woman

I am a strong, independent woman. I advocate for my culture. I stand up for other women like me. I believe that all lives matter. I've overcome a lot of prejudice in my life. I have strong ties to my community. I am very family oriented. I'm raising my children to know that they are equal...

EXAMPLES

Real Black Man

I am Reginald and I am a proud black American. I am very much in love with who I am and though due o many negative things that happen to me die to my race, I still love being a black man I have a loving wife and two amazing kids. I am a christian and a catholic at that...

Real White Man

Born in Iowa, raised in las Vegas, lives in Minneapolis. Only child, not much family. 37 years of age, body feels more like 50. Married for 5 years. Home owner, no kids, steady full time job. Interests are record collecting and Tiki culture. My wife has 2 cats so u guess I have 2 cats...

Fake White Man

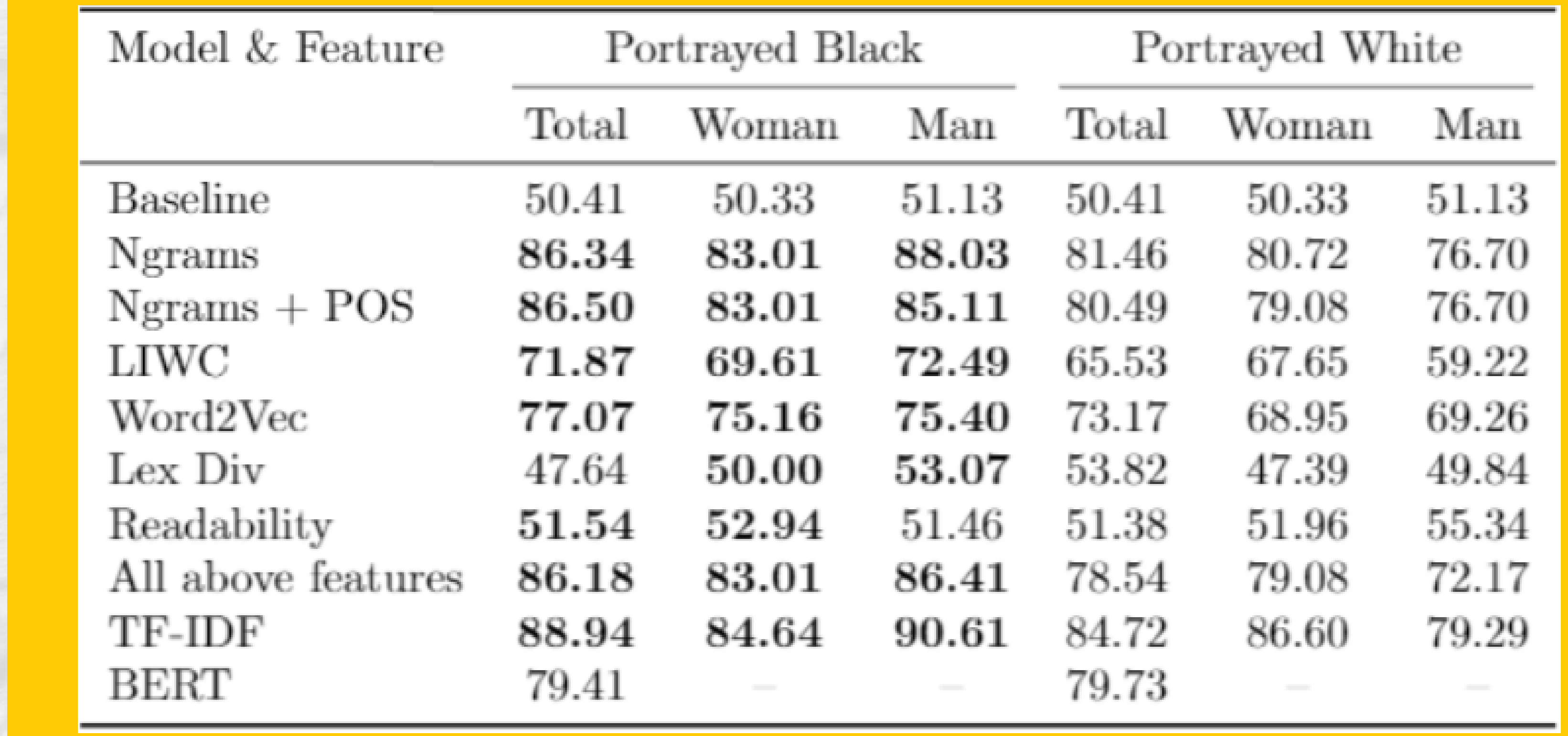
I am Gregory Greene. I am the second born of five kids. My parents hail from the south and are truly loving but also disciplinarians. I was raised to uphold hard work, dedication and trustworthiness. I am married to a beautiful lady with three kids. I am a Methodist as well...

Fake Black Man

Ight, check it. I finna roll up on this couch with my kicks up high. Da f*** the remote at? Who be calling my phone, dats my girl, yo. Lemme spit this out. My girl be hungry, but I'm like naw. I ate and s***. Yo check it, these cats is hungry. Then Simpsons be playing while I trip out and s***...

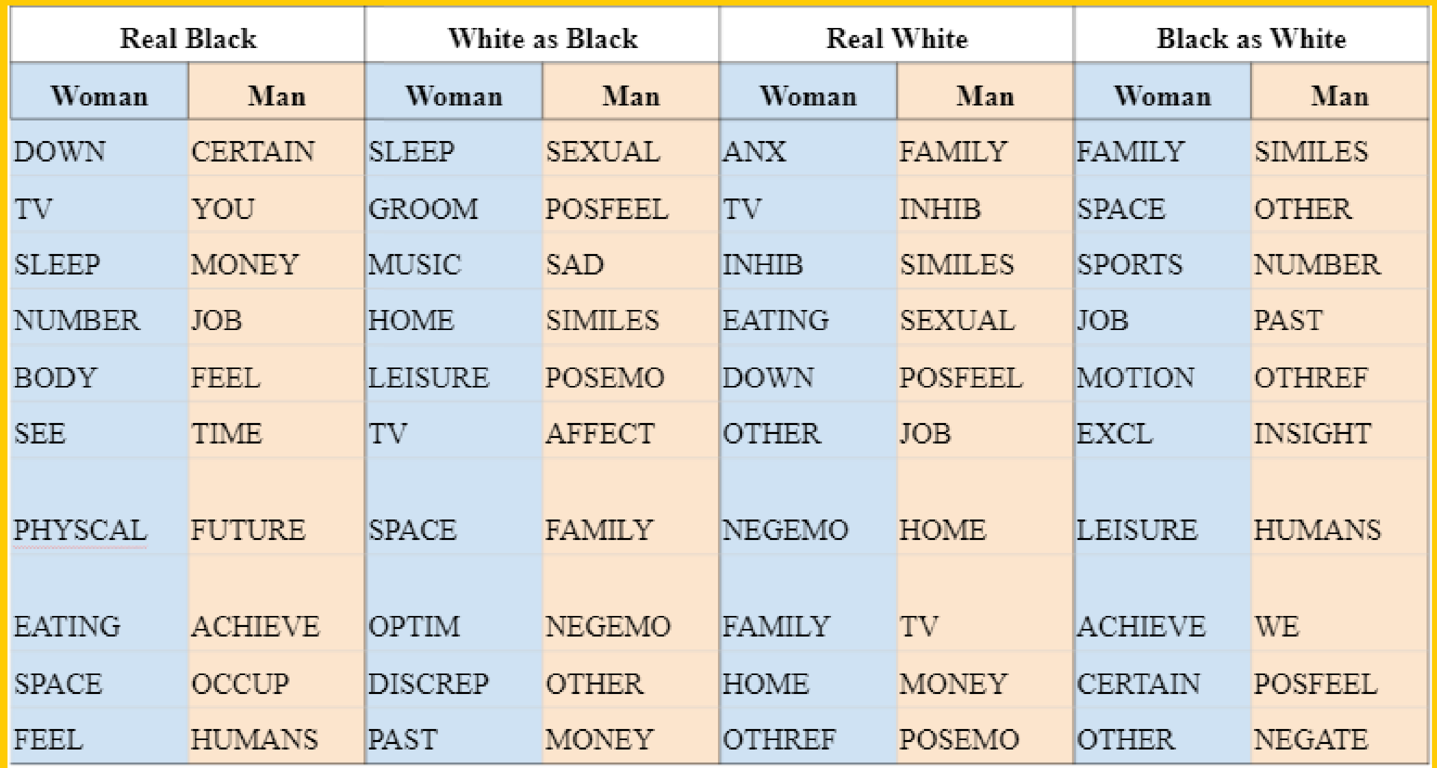
Experiment:
Classification of Racial Identities

- Linear SVM
- Leave one out cross validation
- Features: ngrams, POS tags, LIWC categories, lexical diversity, readability, word2vec (gensim), combined
- TF-IDF feature vectors
- BERT pre-trained embeddings (Bert for sequence classification)
- Predict true racial identity given a response



Model & Feature	Portrayed Black			Portrayed White		
	Total	Woman	Man	Total	Woman	Man
Baseline	50.41	50.33	51.13	50.41	50.33	51.13
Ngrams	**86.34**	**83.01**	**88.03**	81.46	80.72	76.70
Ngrams + POS	**86.50**	**83.01**	**85.11**	80.49	79.08	76.70
LIWC	**71.87**	**69.61**	**72.49**	65.53	67.65	59.22
Word2Vec	**77.07**	**75.16**	**75.40**	73.17	68.95	69.26
Lex Div	47.64	**50.00**	**53.07**	53.82	47.39	49.84
Readability	**51.54**	**52.94**	51.46	51.38	51.96	55.34
All above features	**86.18**	**83.01**	**86.41**	78.54	79.08	72.17
TF-IDF	**88.94**	**84.64**	**90.61**	84.72	86.60	79.29
BERT	79.41	–	–	79.73	–	–

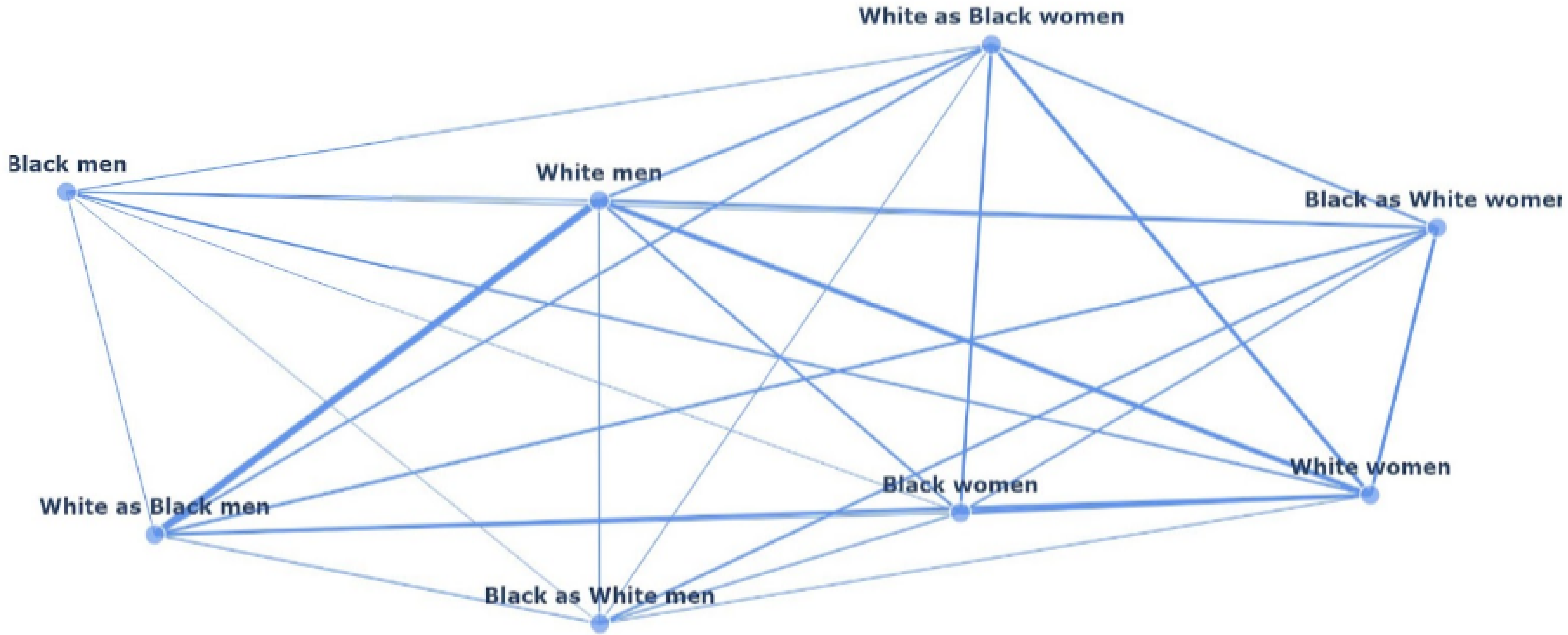
Accuracy values for classification



Real Black		White as Black		Real White		Black as White	
Woman	Man	Woman	Man	Woman	Man	Woman	Man
DOWN	CERTAIN	SLEEP	SEXUAL	ANX	FAMILY	FAMILY	SIMILES
TV	YOU	GROOM	POSFEEL	TV	INHIB	SPACE	OTHER
SLEEP	MONEY	MUSIC	SAD	INHIB	SIMILES	SPORTS	NUMBER
NUMBER	JOB	HOME	SIMILES	EATING	SEXUAL	JOB	PAST
BODY	FEEL	LEISURE	POSEMO	DOWN	POSFEEL	MOTION	OTHREF
SEE	TIME	TV	AFFECT	OTHER	JOB	EXCL	INSIGHT
PHYSCAL	FUTURE	SPACE	FAMILY	NEGEMO	HOME	LEISURE	HUMANS
EATING	ACHIEVE	OPTIM	NEGEMO	FAMILY	TV	ACHIEVE	WE
SPACE	OCCUP	DISCREP	OTHER	HOME	MONEY	CERTAIN	POSFEEL
FEEL	HUMANS	PAST	MONEY	OTHREF	POSEMO	OTHER	NEGATE

Word usage (top LIWC classes)

Word Usage Similarities (LIWC Classes)


White as Black women
Black men
White men
Black as White womer
White women
Black women
White as Black men
Black as White men