2019 Teradata Analytics Challenge Finalist Presentations

TEAM A1-A6

Yenny Yang – Teradata

UNIVERSE

DENVER 2019

teradata.

Format for Session

- Finalist Presentations (Team A1-A6)
- Wrap Up

Analytics Challenge Finalists

A1 - Asia Pacific University Malaysia

A2 - Auburn University

A3 - Lawrenceville High School

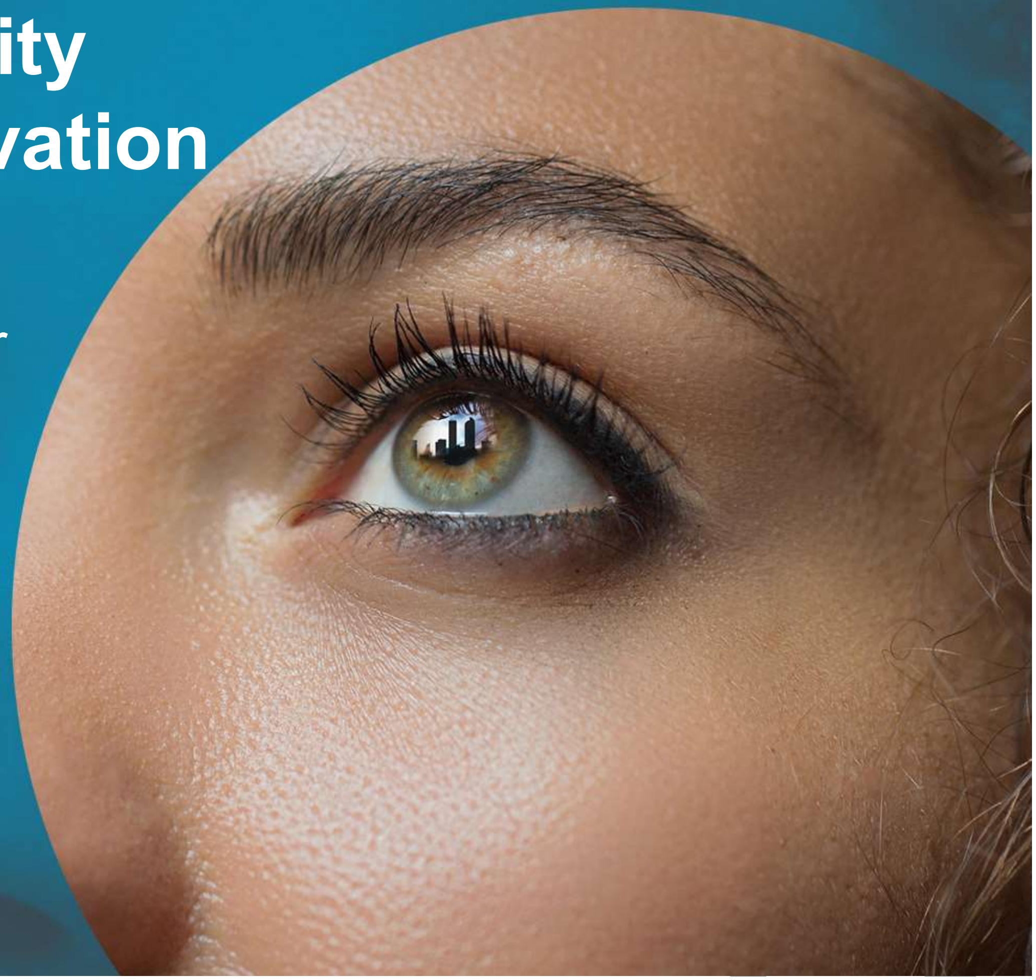
A4- Loyola University Chicago

A5- Southern New Hampshire University

A6- University of Minnesota Duluth

A1- Asia Pacific University of Technology and Innovation Malaysia

Dynamic Pricing in E-commerce Sector



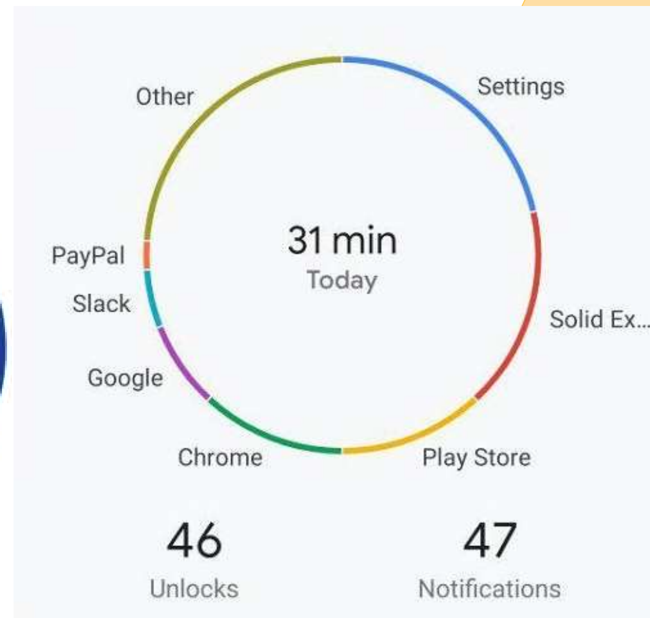
Vijaya Shree Raja Sekaran
October, 2019

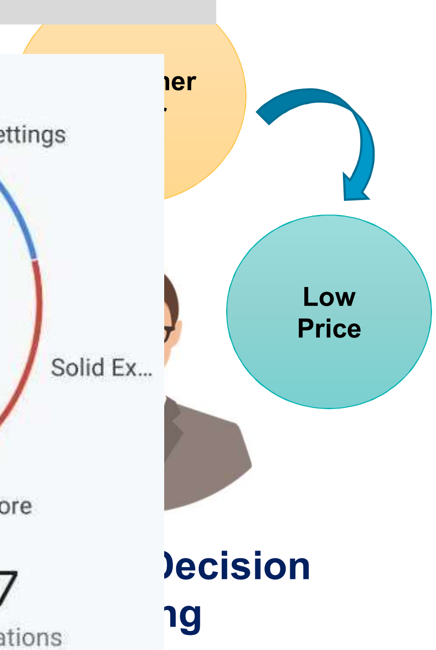
UNIVERSE
DENVER 2019
teradata.

E-Commerce @ Present…!

Data Data Data…

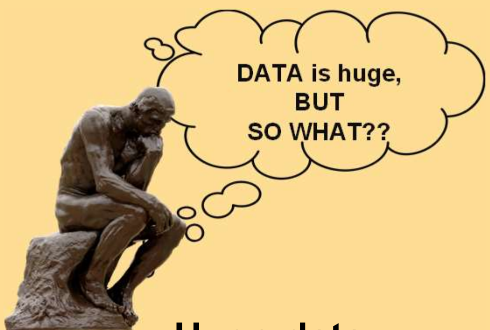
$\frac{40}{80}$ **Mobile Apps**





Problem & Motivation

Problem



Huge data

Price transparency

No significant Financial improvement

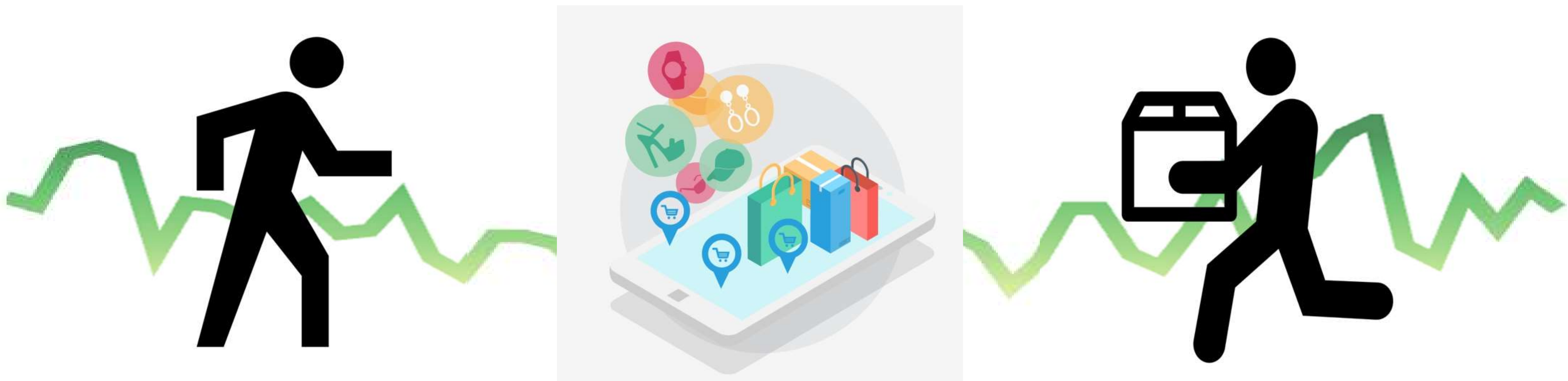
Motivation

Harness the power of ML

Build smart solution

Override business competition

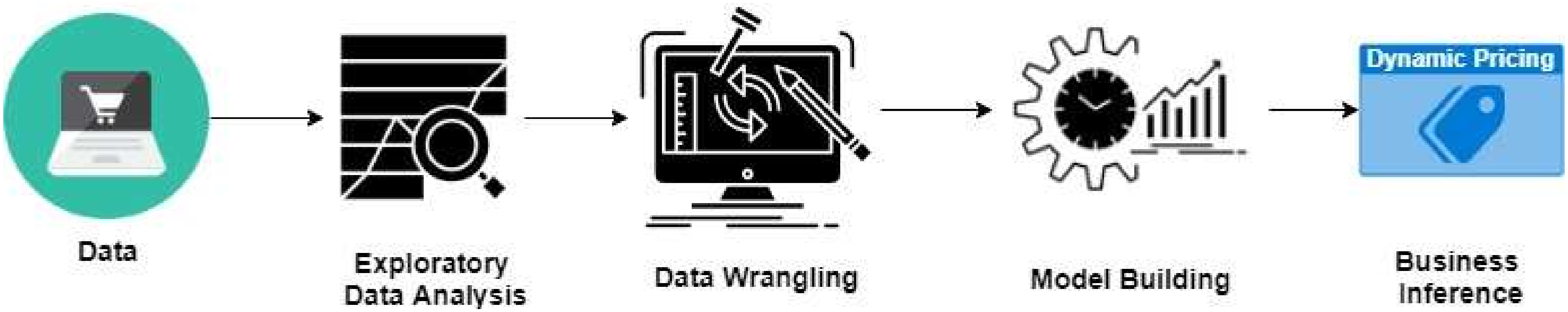
Benefiters of Dynamic Pricing



BUYER　　E-COMMERCE PLATFORMS　　SELLER

Gain customer attention　　**Targeted sales**　　**Increased Profit**

Methodology



Tools & Software

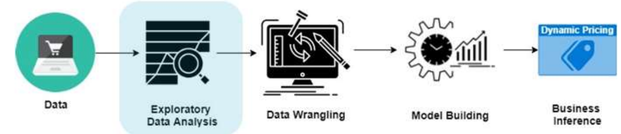
DATAFINITI

tableau SOFTWARE

sas

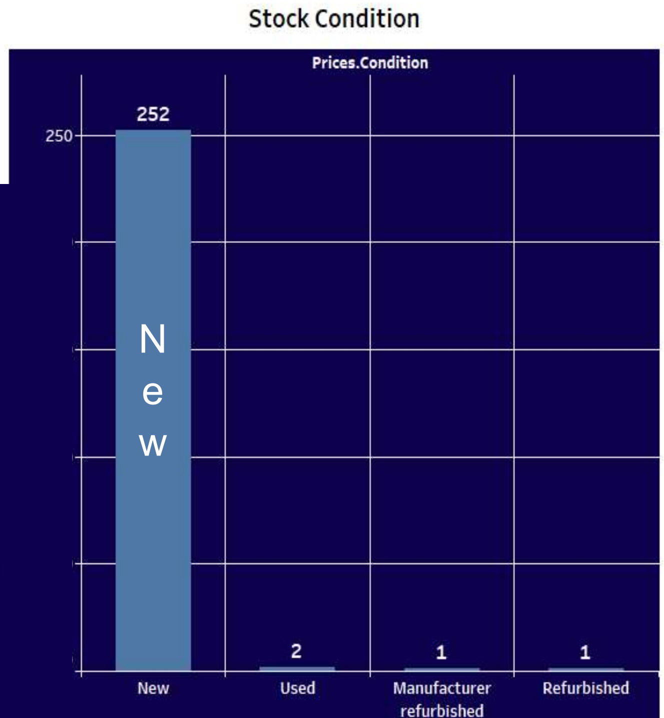
Python 3

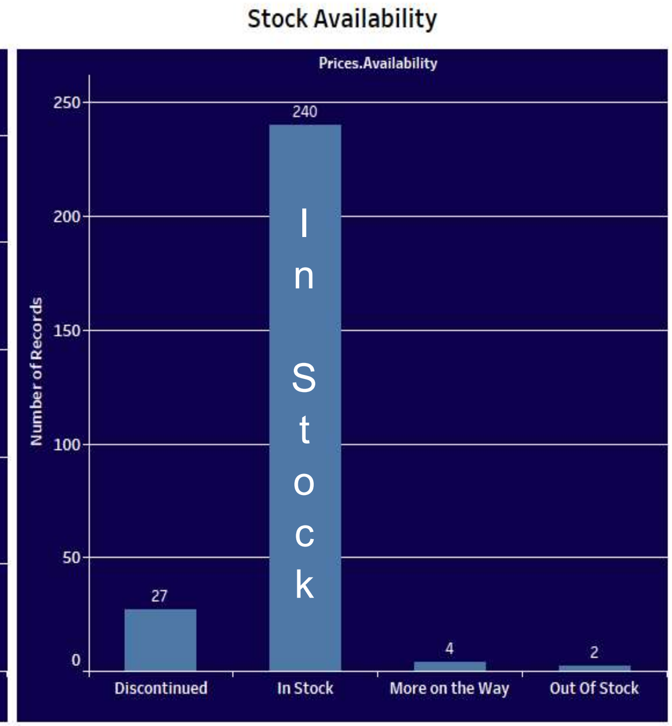
Implication of data analysis

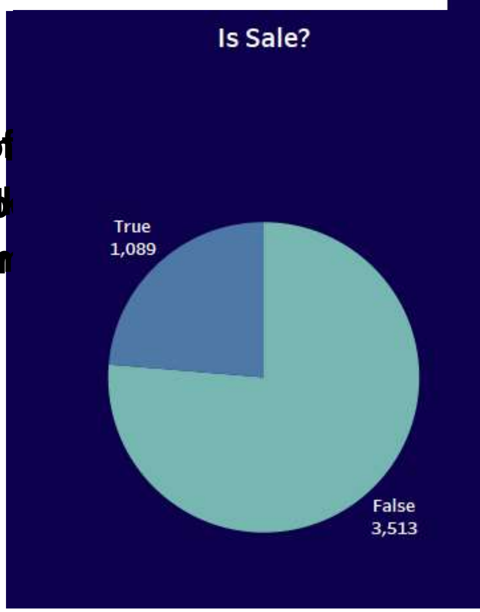


#2

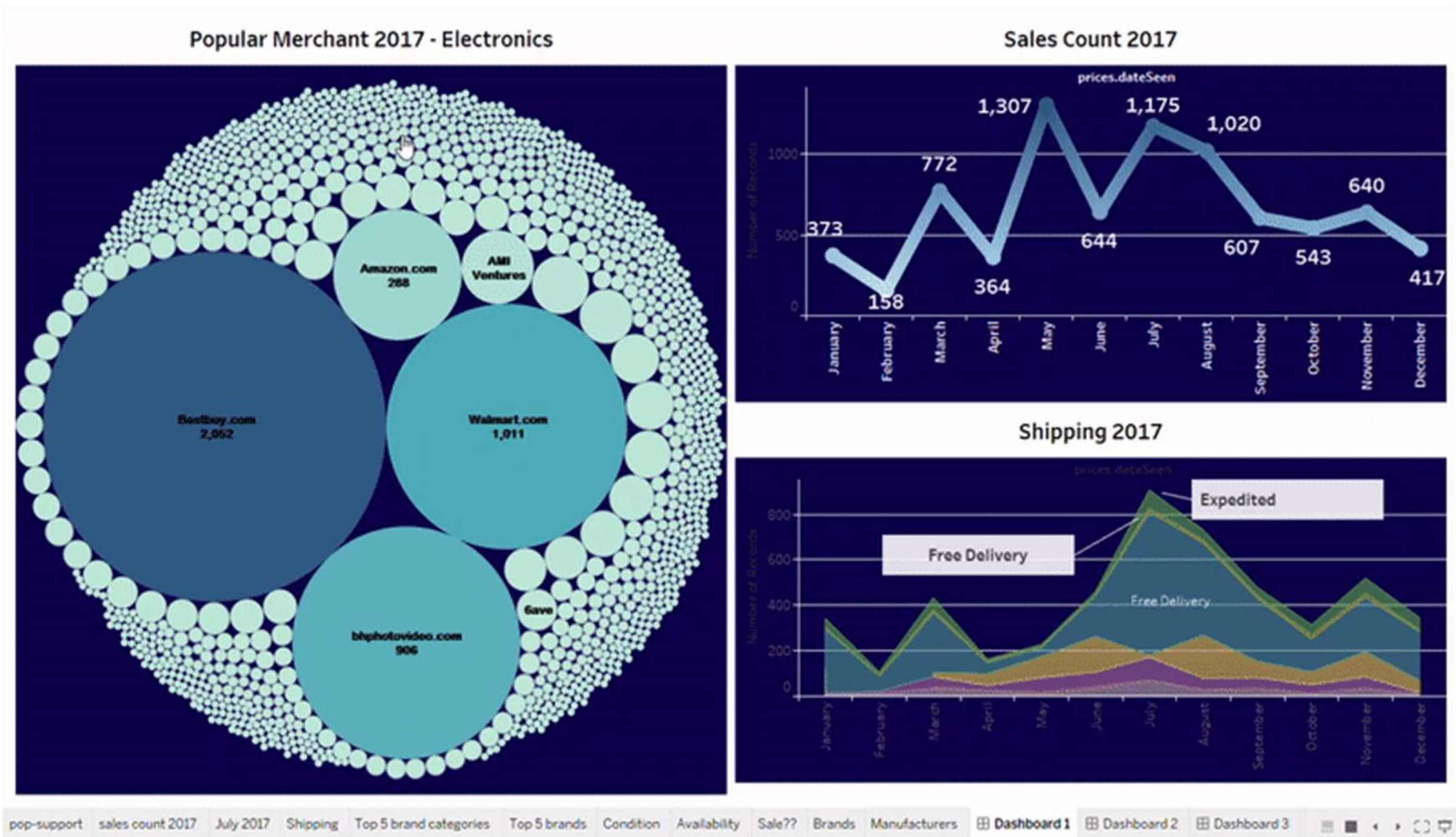
Even though there are no dis
the purchase trend increased







Trends & Patterns in Data



Checkout the interactive dashboard (Tableau)

UNIVERSE
DENVER 2019

8 Significant Factors of this analysis

UNIVERSE DENVER 2019

Analysis on customer behavior & dynamics

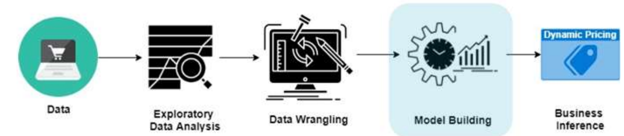
Customer Behavior

Customer Dynamics

Recommendation

Search the products	Enters search term	Track their interest
Add to wishlist	Saves the preference	Match the products in wishlist
Add to cart	Chooses the most probable item to buy	Send the notification of the lowest prices from dynamic pricing strategy
Purchase	Place the order	Add the trait to their dynamics

Recommendation

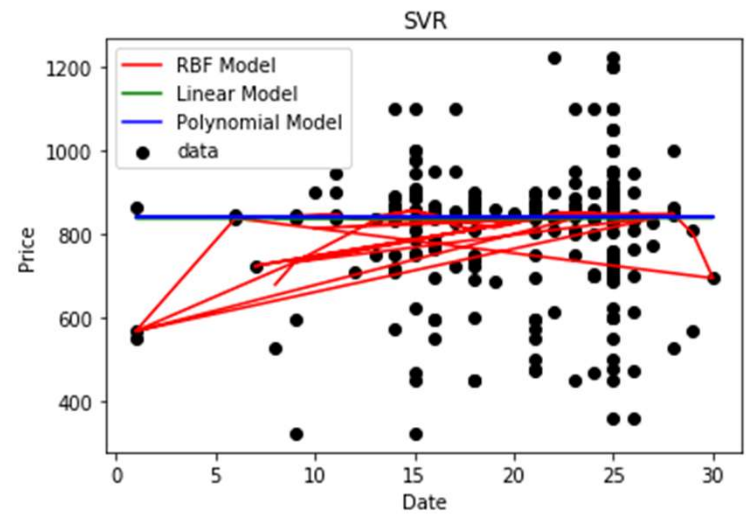


Implementation of predictive ML model

- ✓ Features: Date Seen & Previous price
- ✓ Product: MacBook2017
- ✓ Technique: Support vector machine
- ✓ Kernels: Radial basis Function, Linear, Polynomial

RBF	$Y(Xi,Xj) = \exp(-\gamma\lvert Xi - Xj\rvert^2)$
Linear	$Y=mx+c$
Polynomial	$Y(X_i, X_j) = (X_i \cdot X_j + 1)^d$

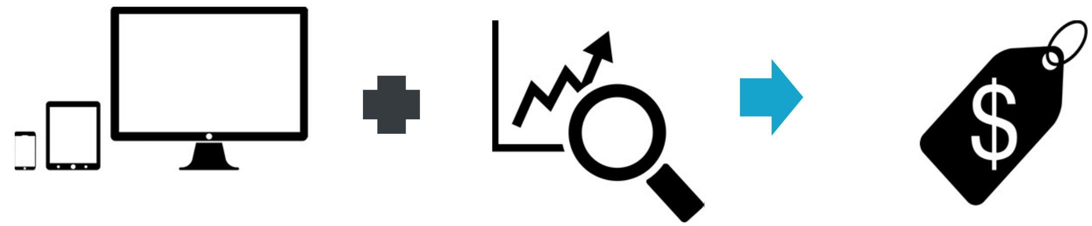
Output of ML model




```
print(predicted_price)
(850.0897439618172, 839.9500000004048, 840.0900000197965)
```

Dynamic pricing

This strategy can highly optimize the profit !



USERS DIGITAL BEHAVIOUR

DATA ANALYTICS

DYNAMIC PRICING

The way forward..

- ✓ Deployment of this model in CL ne.
- ✓ Real-time analytics and price ic w ew features
- ✓ Plug-in development

Thank you

For Demo of the model, Credits & Acknowledgement – Please visit:
Http://MyAnalyticsCompetition.online

Mentor
Mr. Raheem Mafas

Presenter
Vijaya Shree Raja Sekaran

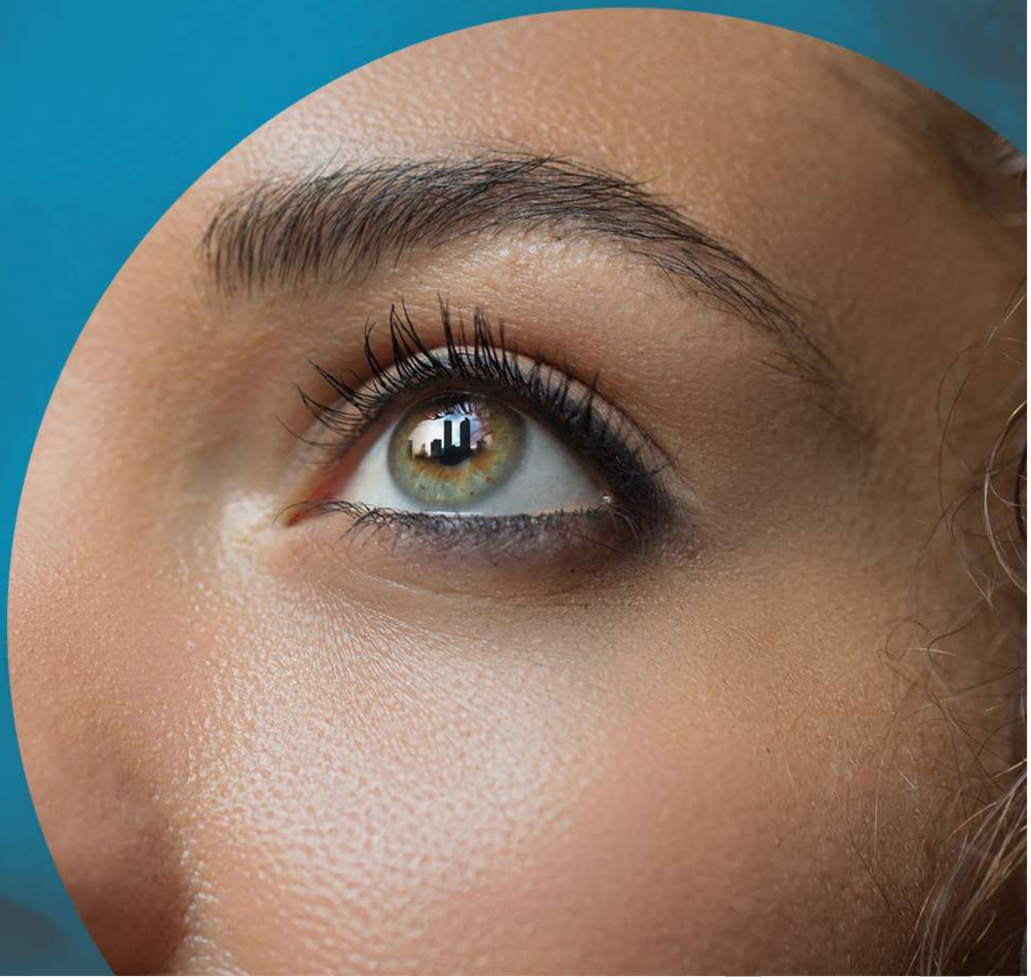
UNIVERSE
DENVER 2019

A2 – Auburn University USA

The Government Shutdown (2018-2019) as seen through Twitter

Maia MacDonald
Lauren Fuller, Avery Lawrence

October, 2019



UNIVERSE
DENVER 2019
teradata.

Introduction

- Between December 22, 2018 and January 25, 2019, the United States government shutdown
  - Longest shutdown in American history
- Began over funding for a wall along the southern US border
- Impacted 800,000 Americans who were working without pay during the shutdown
- Interested in Twitter sentiment relating to furloughed federal employees, management of TSA and air traffic control, and destruction of National Parks

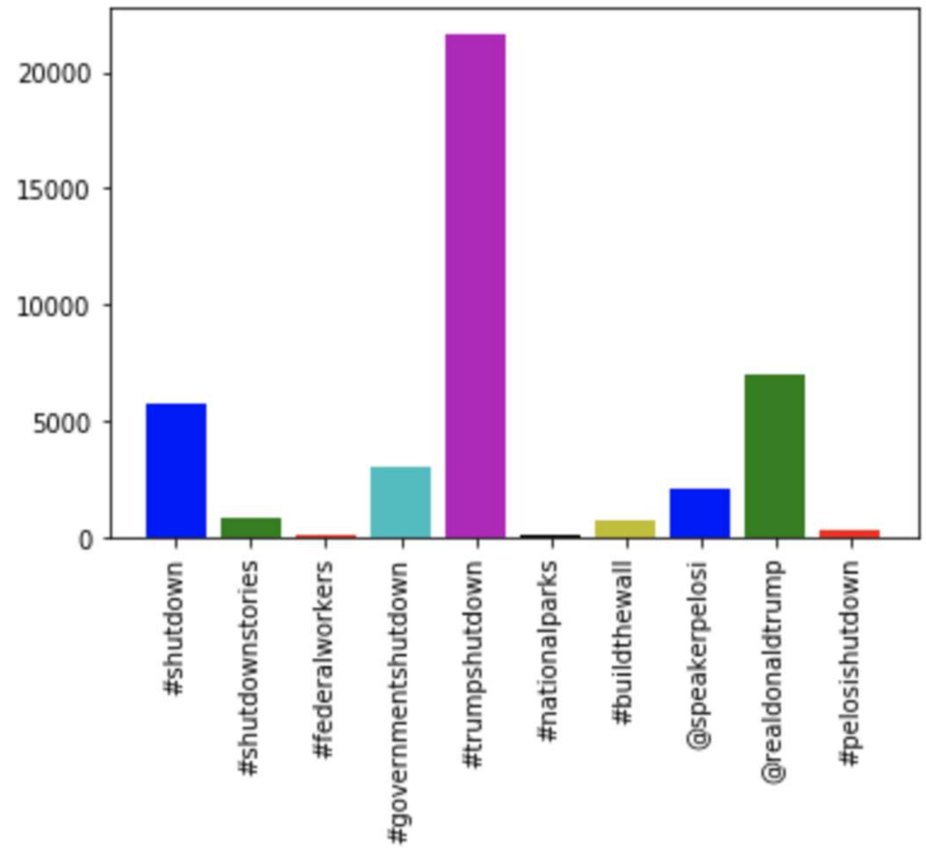
Topic Research & Analytics

- Using Amazon Web Service's Elastic MapReduce (EMR), we queried the listed hashtags from 3 months of Twitter data
  - Collected between December 3, 2018, and February 4, 2019
  - Generated roughly 100,000 tweets
- The data created a considerable amount of noise, so we cut the list down to only **#shutdown** during our analysis
  - The change reduced the number of tweets to 39,262

#shutdown
#governmentshutdown
#endtheshutdown
#tsa
#shutdownstories
#federalemployees
#federalworkers
#scienceshutdown
#airtrafficcontrol
#nationalparks
#coastguard
#nasa

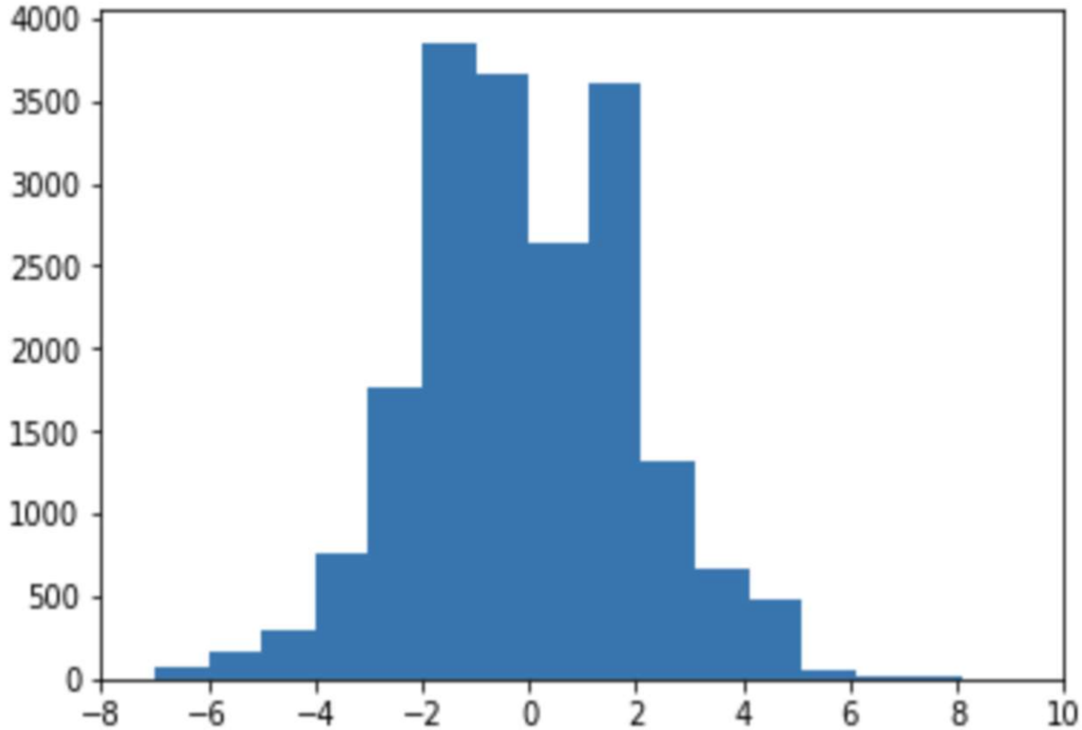
Hashtag Frequency

- Frequency of hashtags and usernames relating to politics:
  - #trumpshutdown – 21,528
  - @realdonaldtrump – 6,981
  - @speakerpelosi – 2,041
  - #buildthewall – 711
- Frequency of hashtags relating to the effects of the shutdown:
  - #shutdownstories – 823
  - #federalworkers – 58
  - #nationalparks – 35



UNIVERSE
DENVER 2019

Aggregate Sentiment



Sentiment Scores for each Hashtag & Username

Fig. 1: Average Raw Sentiment Scores

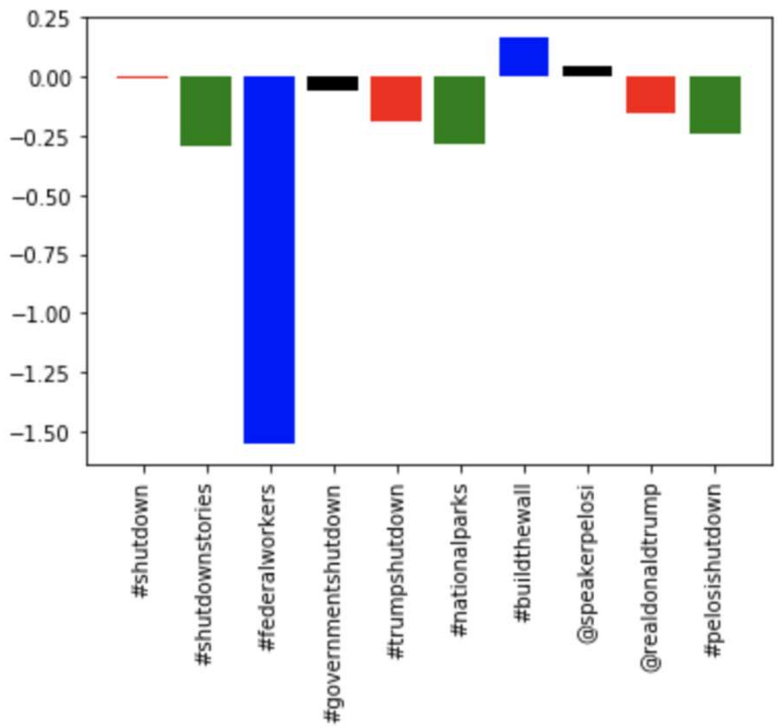
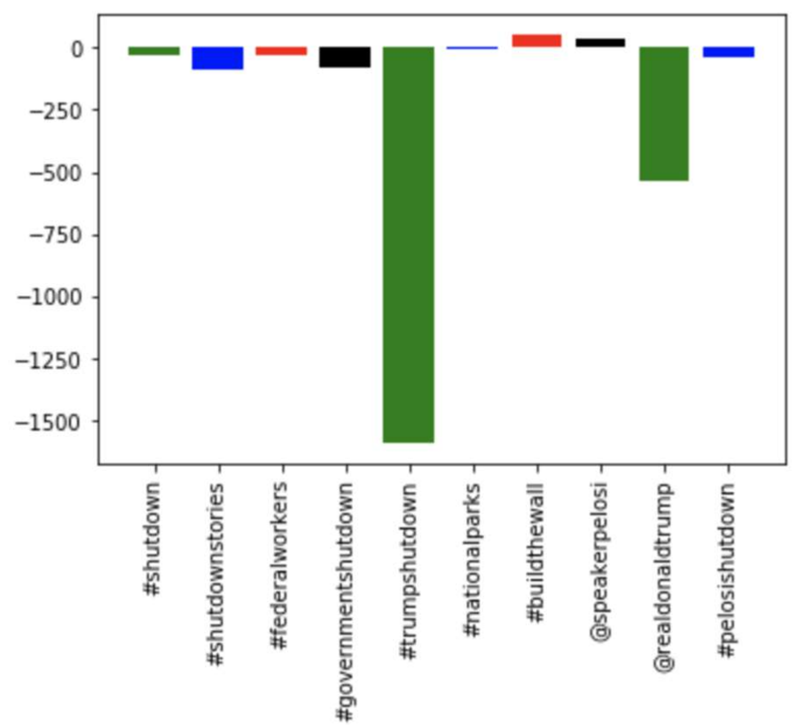


Fig. 2: Aggregate Raw Sentiment Scores



UNIVERSE
DENVER 2019

Proportion of Positive & Negative Sentiment Scores

Fig.1: #federalworkers

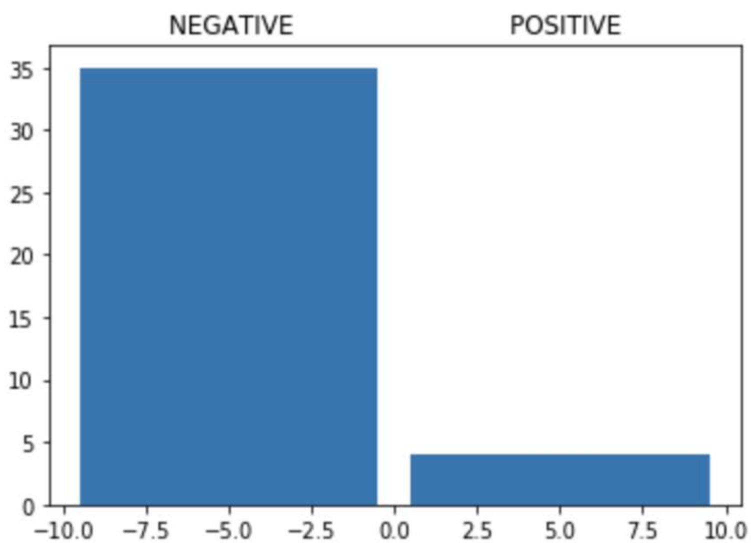
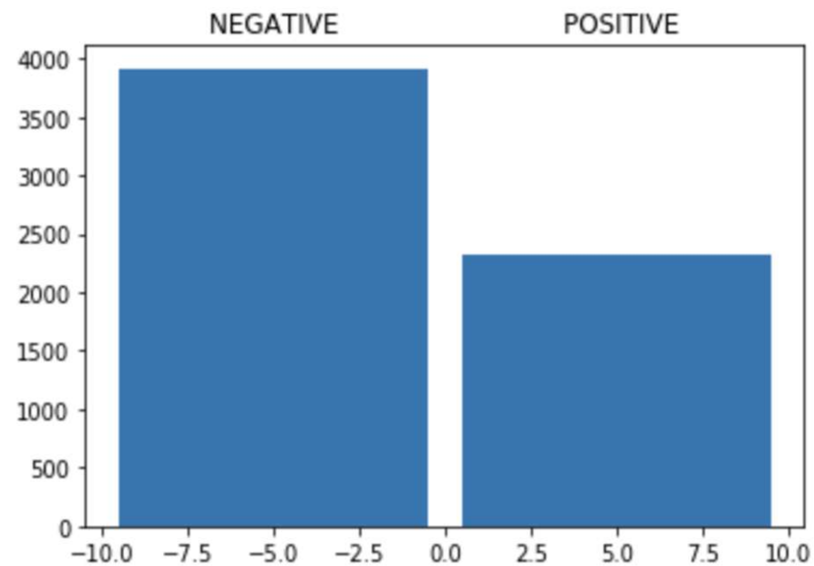
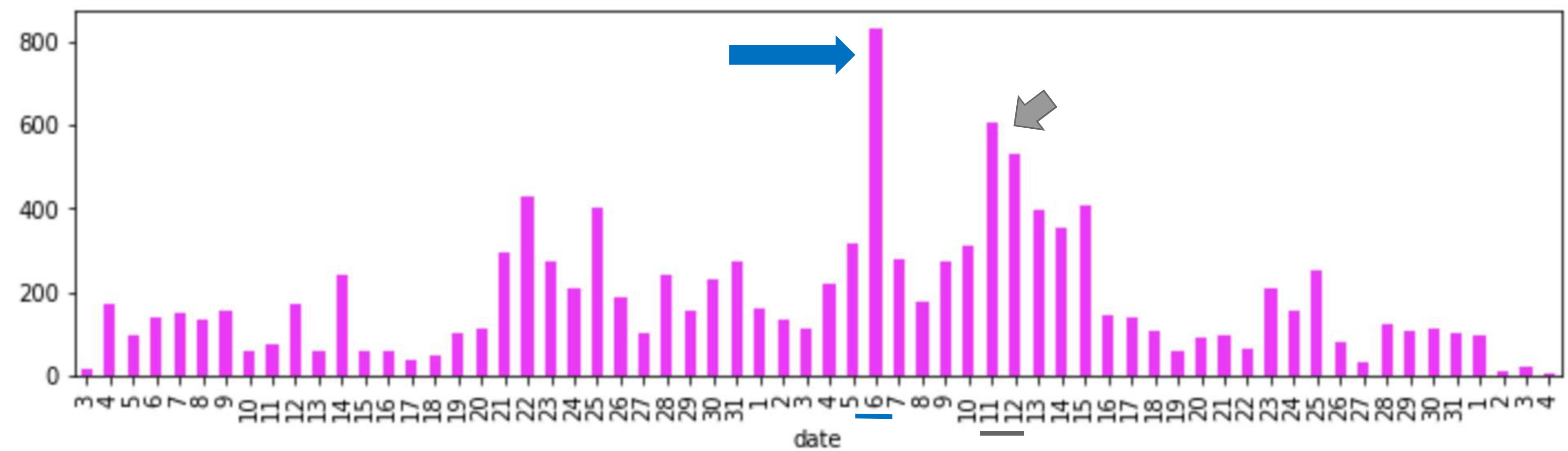


Fig. 2: #trumpshutdown



Frequency of Tweets on Queried Dates



Highest Frequency Hashtags

Fig. 1: January 6, 2019

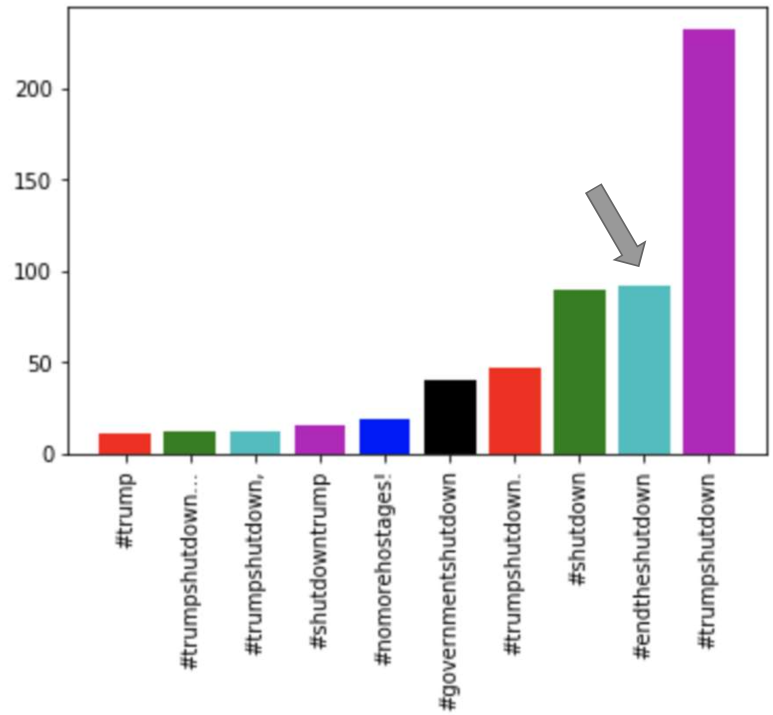
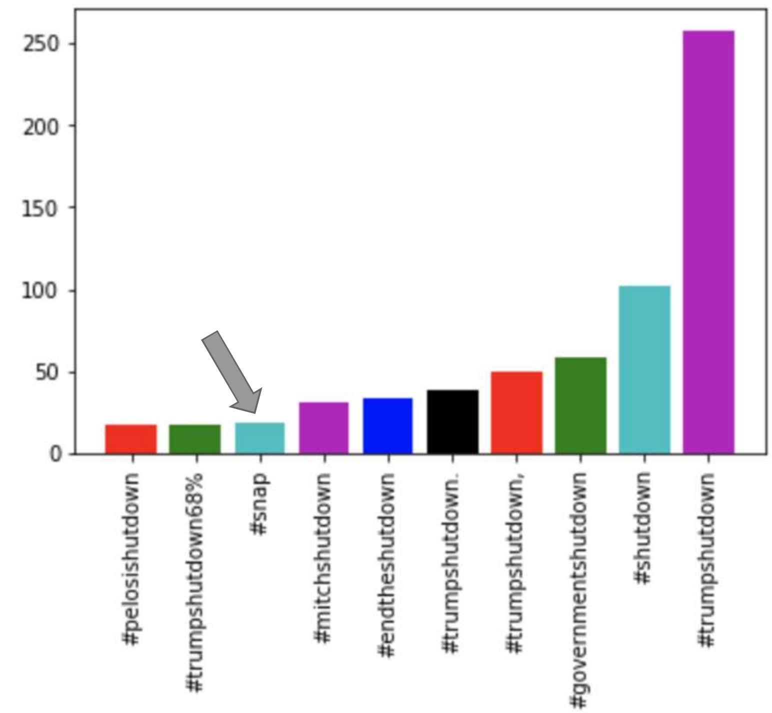
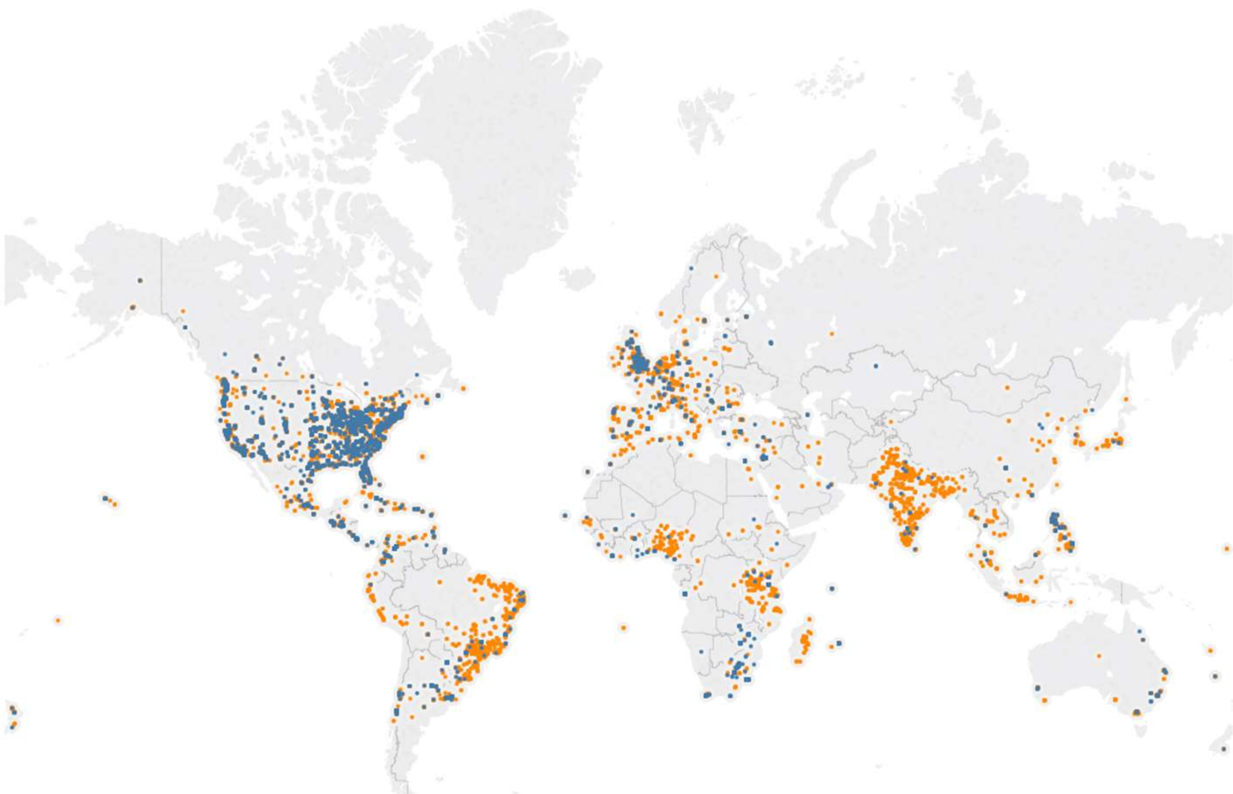


Fig. 2: January 12, 2019



UNIVERSE
DENVER 2019

Trump vs. Non-Trump Tweets



Tweets that contain 'trump'

Tweets that do not contain 'trump'

27,471 Trump Tweets
11,791 Non-Trump Tweets

UNIVERSE
DENVER 2019

Conclusion

- We expected people to express concerns with their jobs and paying their bills
- Instead, people direct their anger towards political figures
  - Because of this, analytics techniques become difficult to use for assessing socioeconomic impacts of government actions
- Twitter should be avoided for business analytics purposes when the issue is related to politics
  - Other forms of social media may be more useful

Thank you.

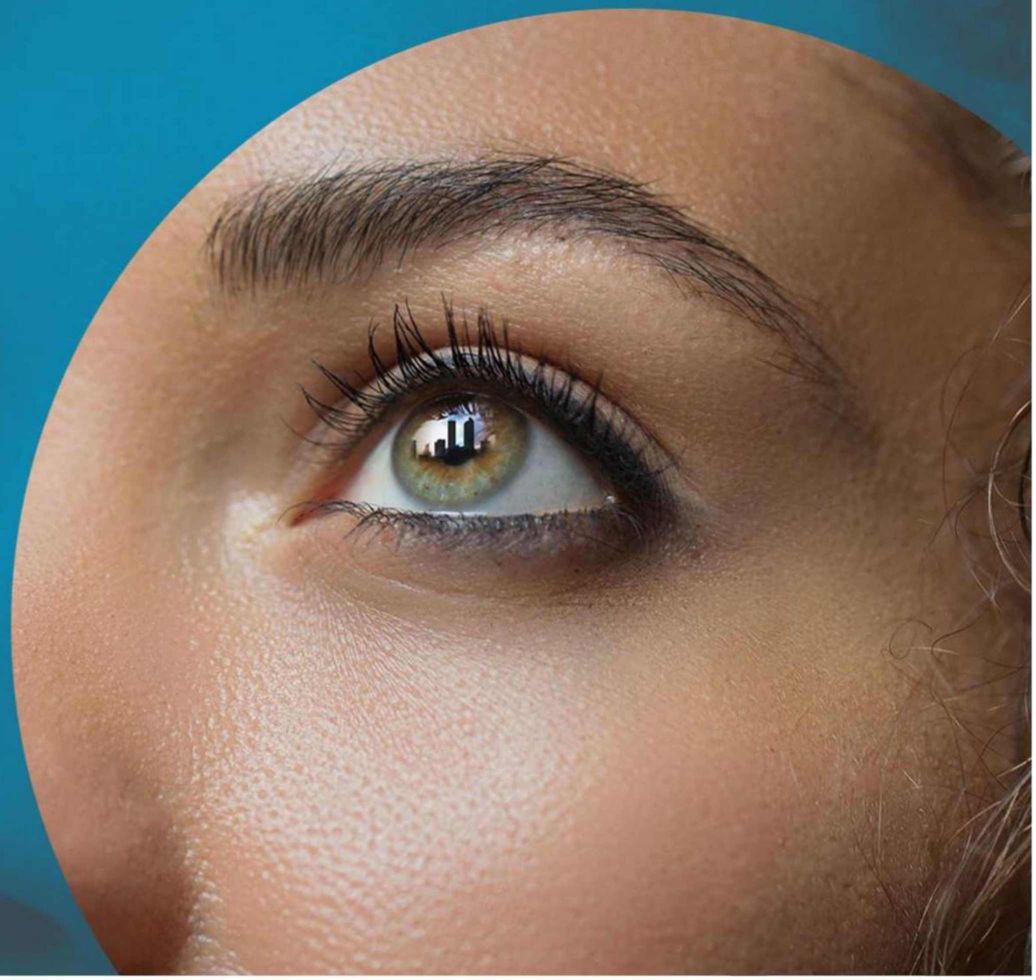
UNIVERSE
DENVER 2019

A3- Lawrenceville School (HS) USA

Is Legalisation of Cannabis for Adults Harming Teenagers?

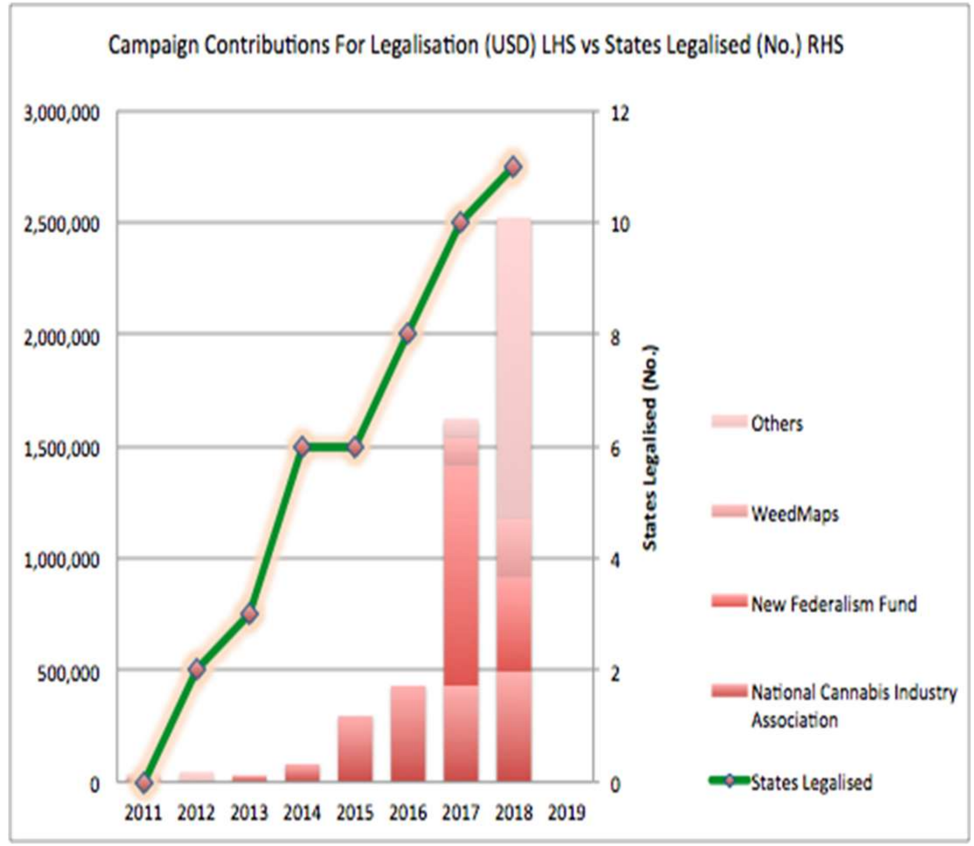
Johann Lee
Alen Alshinbayev
October, 2019

UNIVERSE
DENVER 2019
teradata.



Our Motivation

Is Cannabis Legalisation for Adults Harming Teenagers?



Medical Research shows Major Health Risks
To Teenagers:

- Poorer education/employment achievements
- Impeded cognitive abilities
- Increased car accidents
- Increased addiction to hard drugs in later life

Note: "Adverse Health Effects of Marijuana Use" by Volkow et al. published in New England Journal of Medicine (2014) and many other research papers

Data and Tools - What did we look at?

Clean Datasets → **Apply Tools** → **Analyse for Correlation and anomalies**
- Scatterplots
- Box plots

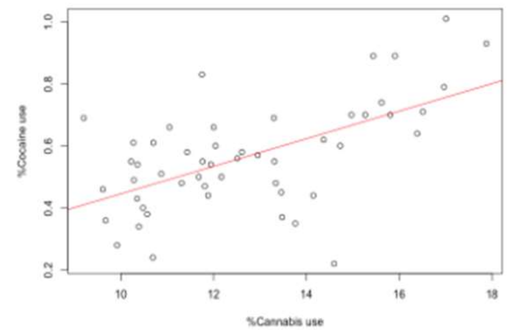
United States® Census Bureau

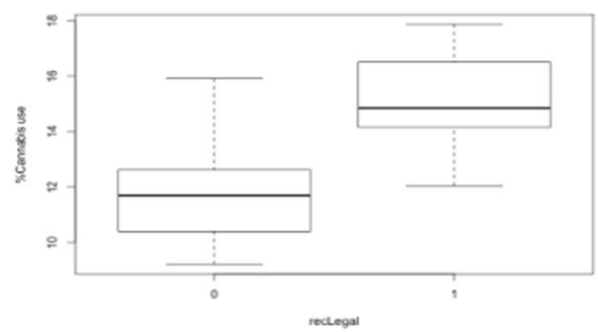
SAMHSA Substance Abuse and Mental Health Services Administration

R Studio

tableau software

Excel





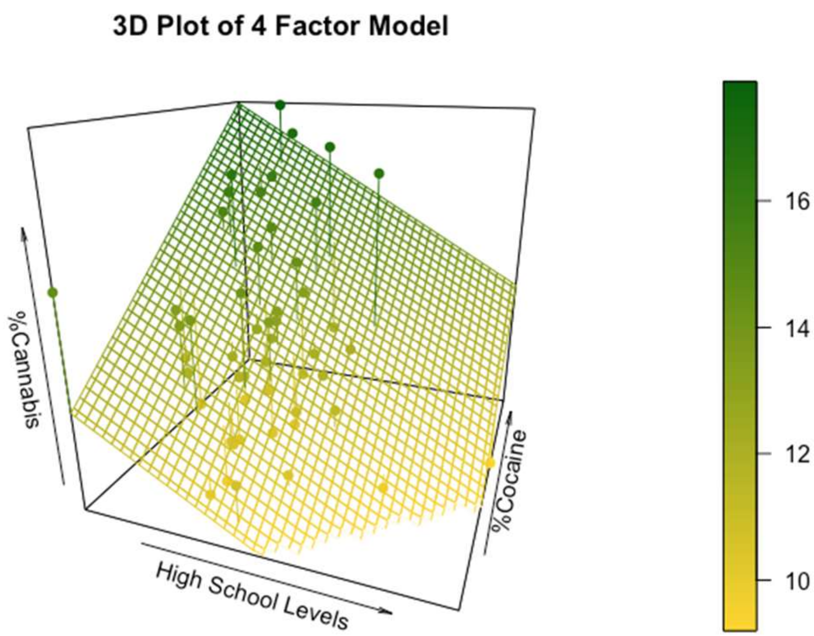
UNIVERSE DENVER 2019

Results: What did we find?

Best Model - 4 factors (legal status, cocaine usage, high school levels, depression level)

(F-statistic 26.35 on 46 Degrees of Freedom, R-squared 0.70, overall p-value of < 0.0001)

Multiple Linear Regression

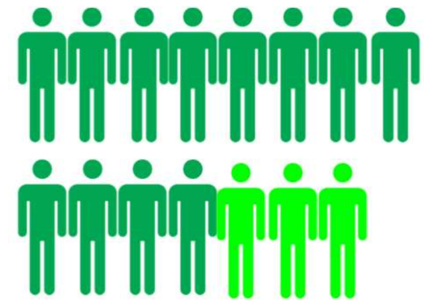


$$y_i = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots \beta_n x_n$$

When we legalize:

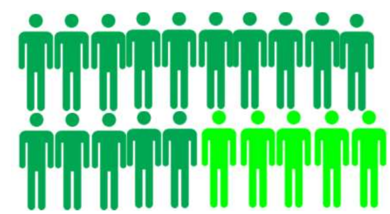
For every 100,000 teenagers

Teenage Cannabis Use: +22%



Teenage Cocaine Use: +21%

High School Dropout rates: +30%

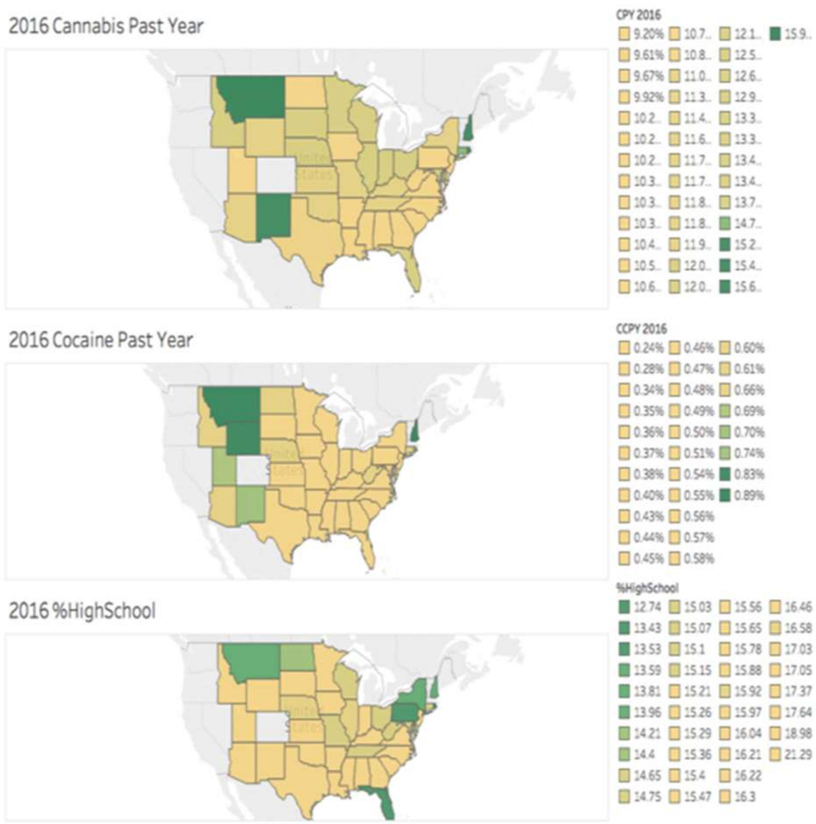


*Each person represents 1000, except cocaine (100)

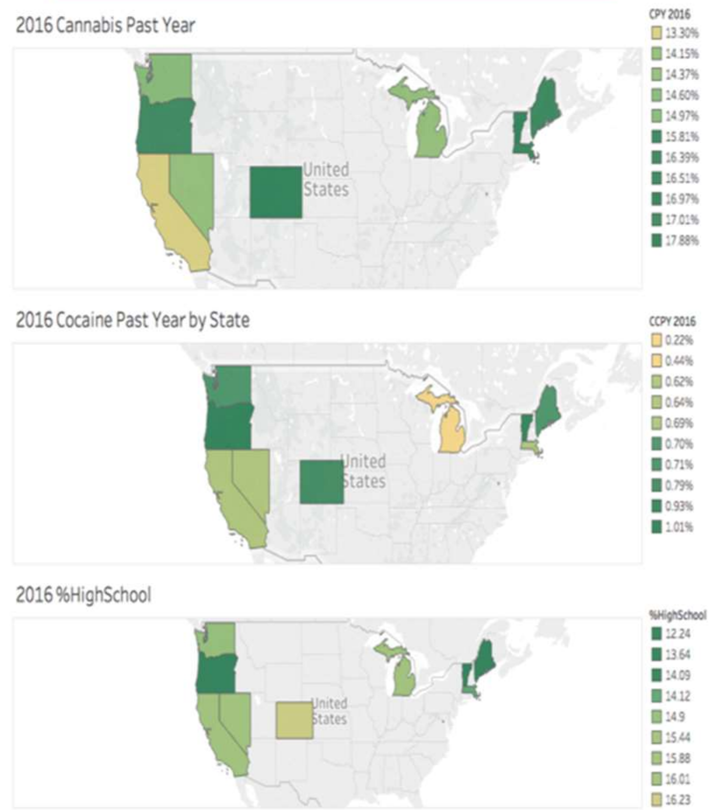
UNIVERSE
DENVER 2019

Results:Visualisation

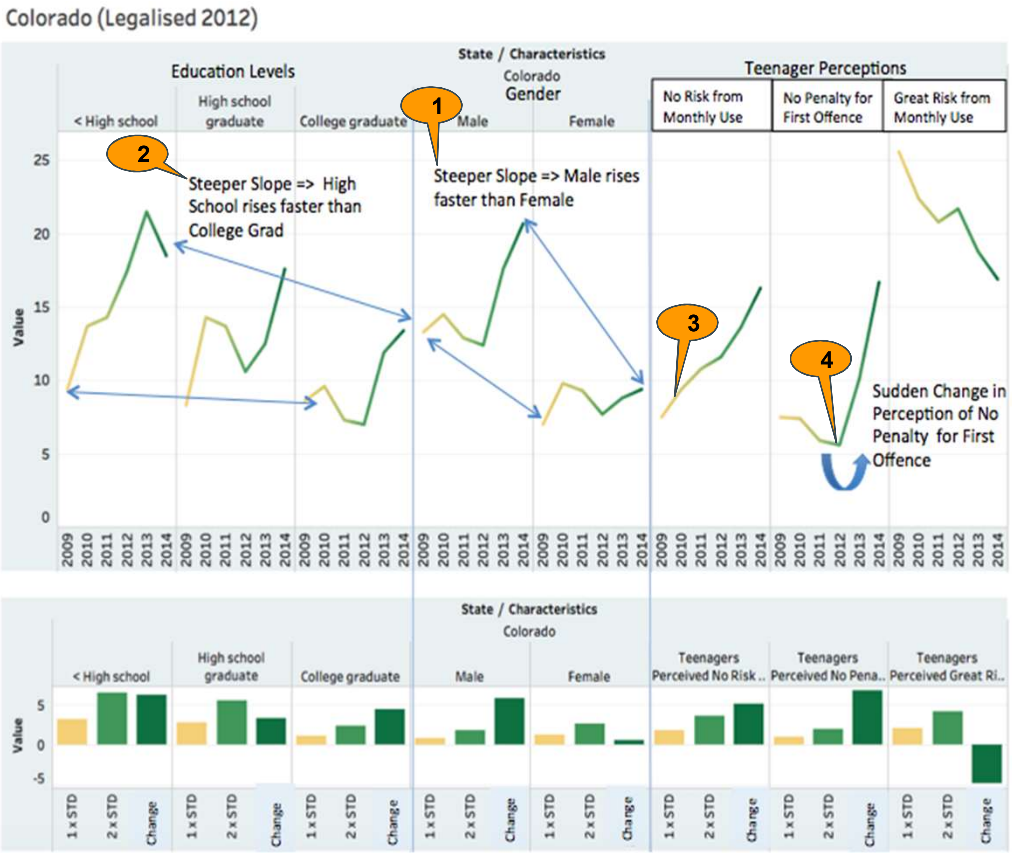
Illegal States (excluding Hawaii)

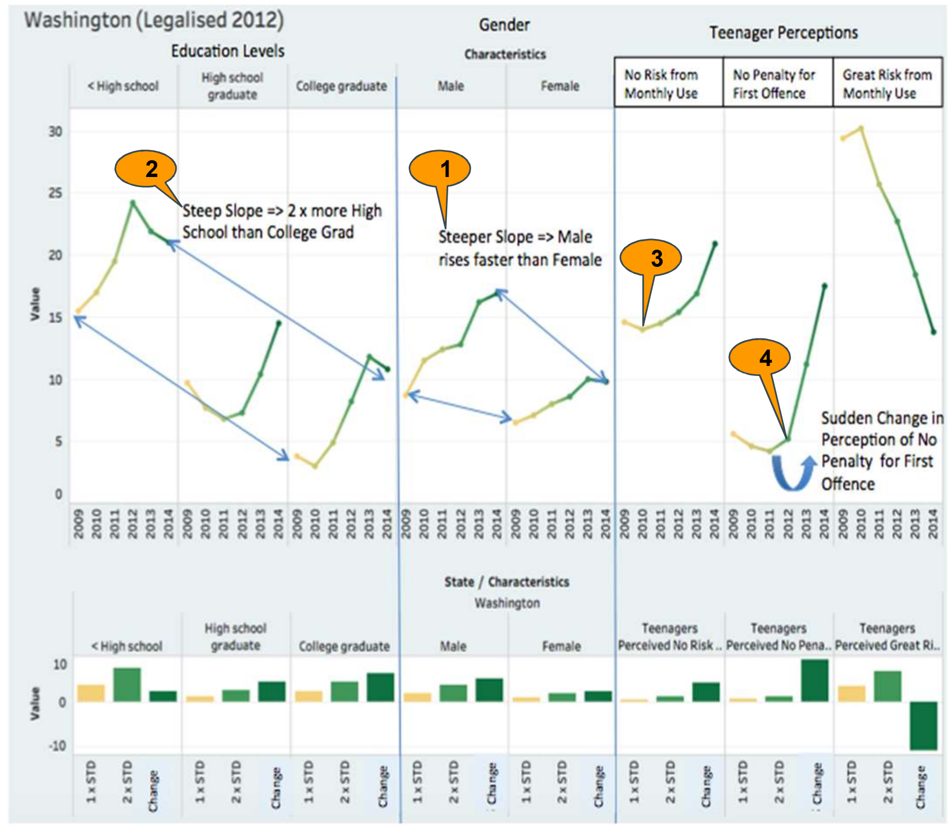


Legal States (excluding



Results: Changes in CO and WA over 5 years





Recommendations

Teenage Males with Poor Education
- **Doubling usage**
- **Misconceptions of:**
  - **no health risks**
  - **no legal penalties**

VS

Teenage Tobacco Use:
- **Falling steadily from 12% (2008) to 6% (2016)**

- Increase
  - Taxes
  - Education
  - Legal Enforcement
- Support from influencers

Thank you.

A4- Loyola University Chicago USA

What Bike-sharing Data Reveals about Chicago's Socio-Economic Landscape

Nick Damato and Grace Sperr

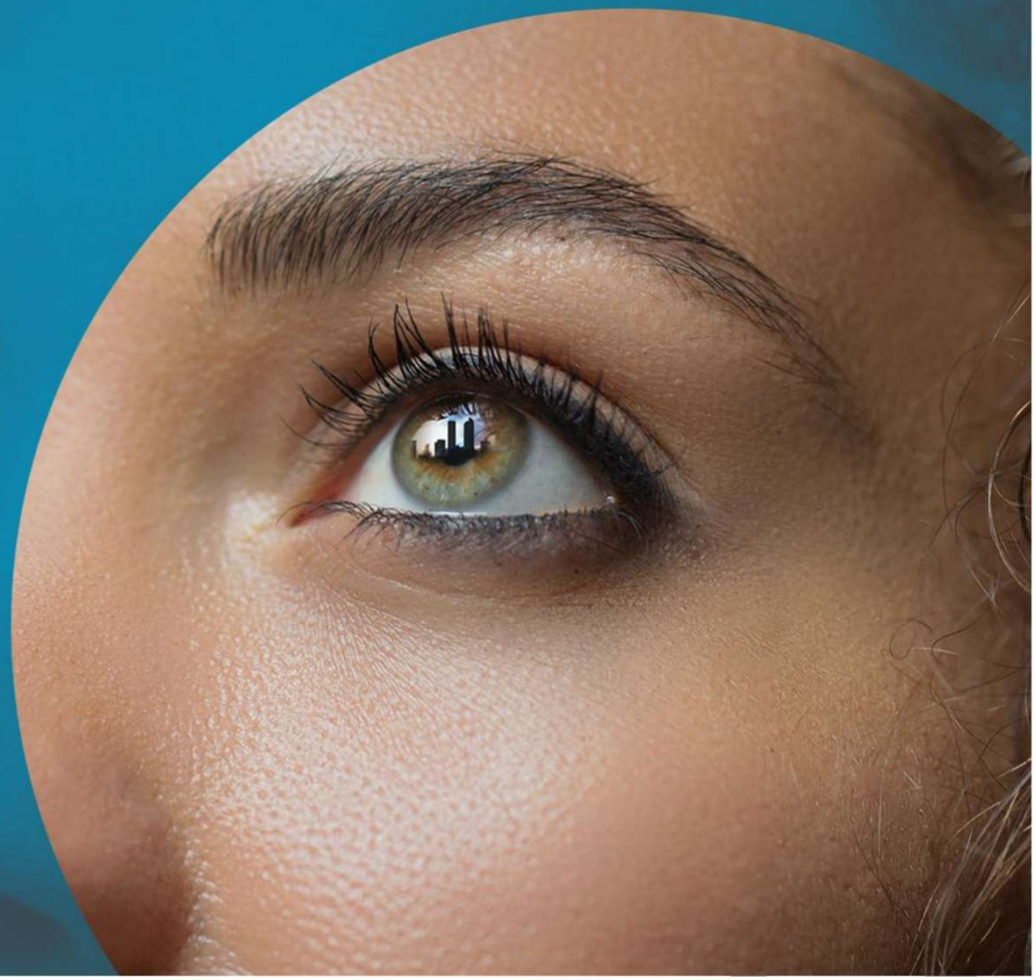
Allison Heithoff, Varsha Kalangari, and Eric Stepanovic

October, 2019

UNIVERSE

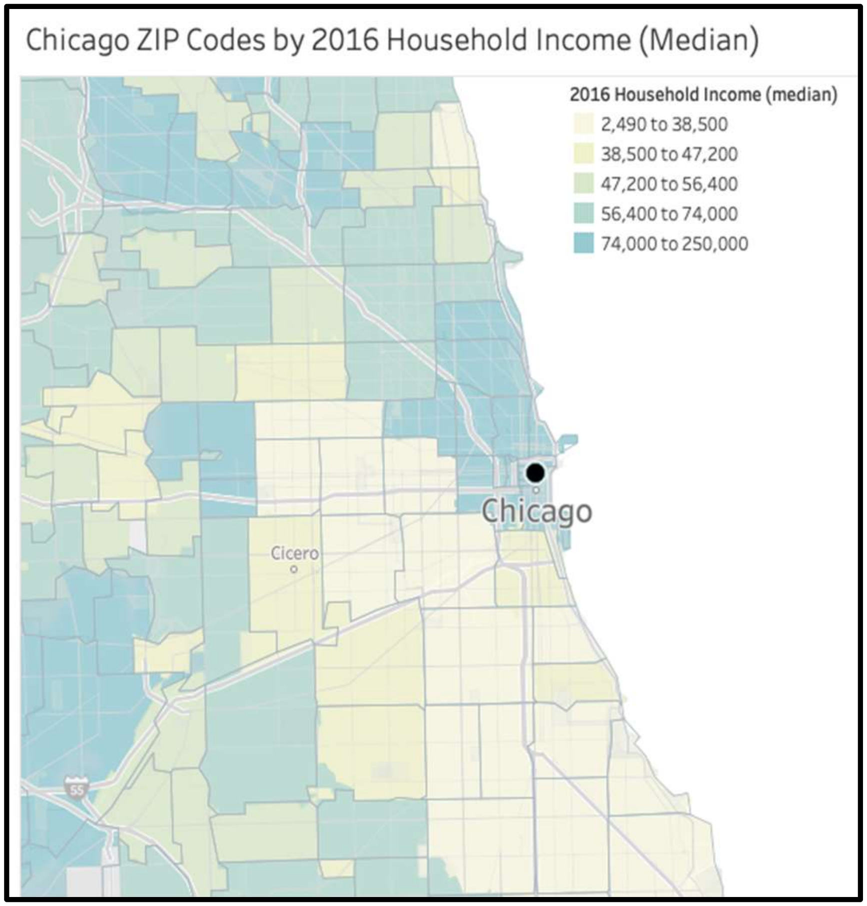
DENVER 2019

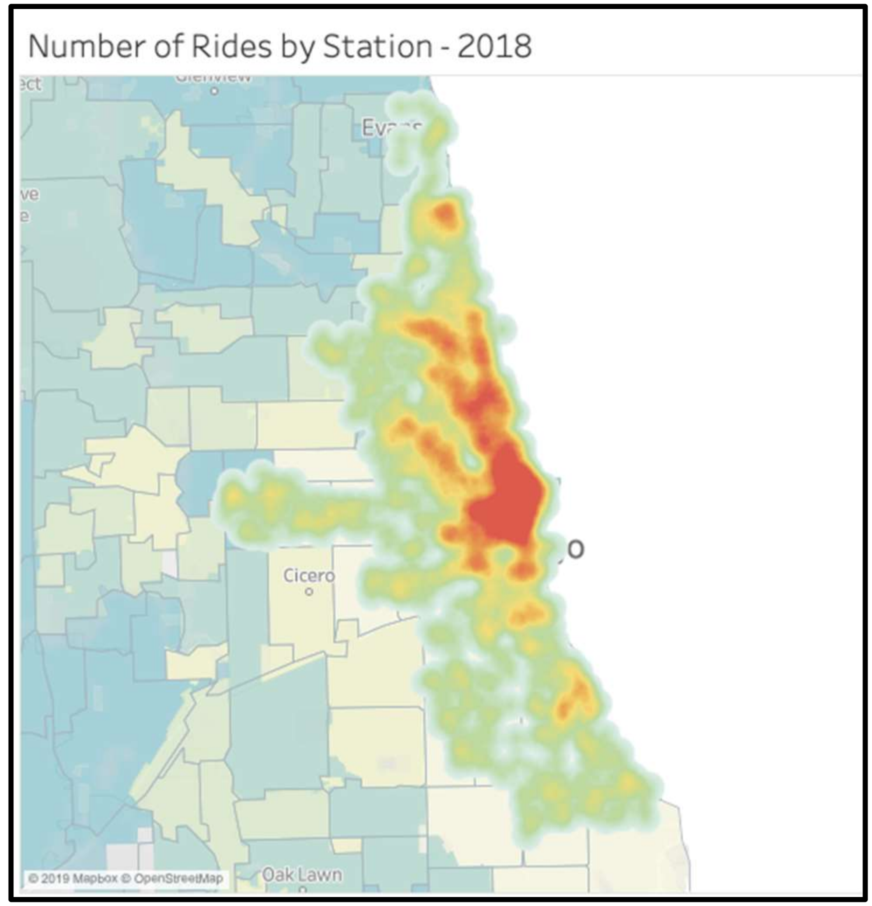
teradata.



Problem: Bike-sharing ride trends reflect the same inequality as Chicago’s socio-economic landscape

In 2018 majority of rides occurred in affluent areas of the city

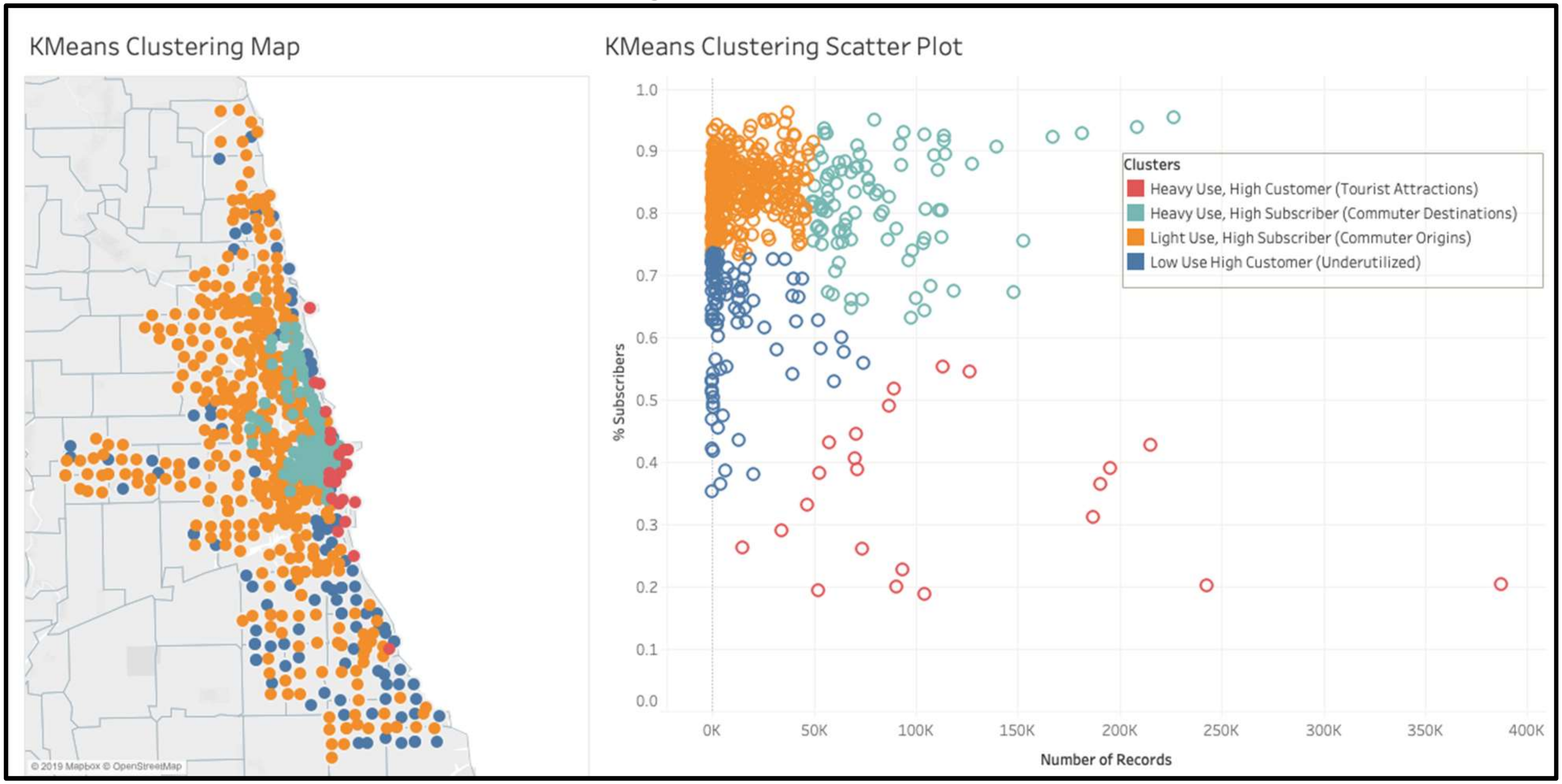




UNIVERSE
DENVER 2019

Station Clusters

When clustered using K-Means clustering in R, based on non-location attributes, the 4 groups mirrored the same socioeconomic landscape as Chicago income data



UNIVERSE DENVER 2019

Motivation: Understand the barriers to bike-sharing in low-income areas and explore how the city, organizations, and local communities can overcome these barriers to create a more equitable city through bike sharing

Data Sources

DIVVY

2013 - 2018 Trip Dataset
17 million records
9 dimensions

United States Census Bureau

CHICAGO DATA PORTAL

Analytics Tools

tableau

R Studio

PostgreSQL

Approach: Brainstormed potential barriers, before choosing 3 focus areas to explore

Crime

Resource Availability

Labor

Gender

Time Limit

Demographics

Distance

Cost

Education

Public Health

Infrastructure

Resources

Hypothesis: Low-Income areas have less stations, therefore increasing stations will increase ridership

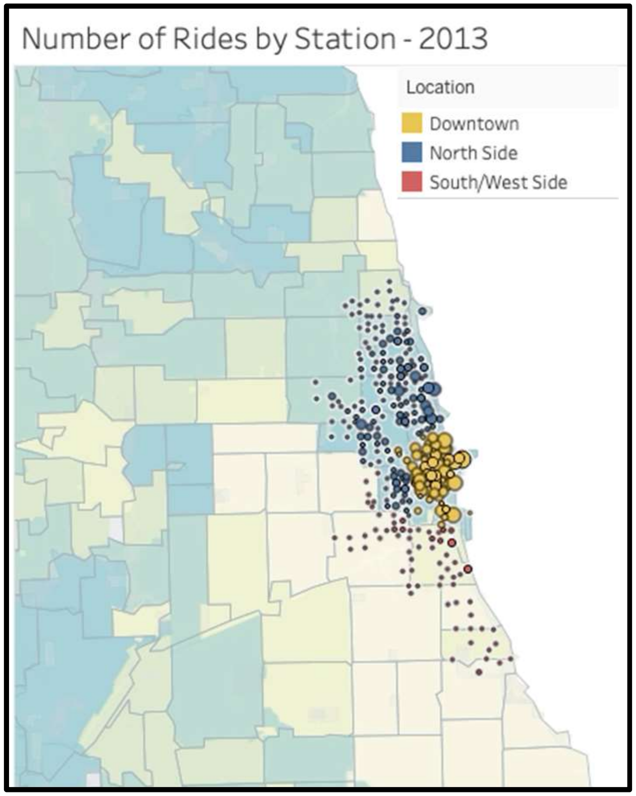


Figure 1 (video)

Figure 1 shows Divvy's explosive growth over the past 5 years, both geographically and in usage

Figure 2 shows that as the number of stations increased in South Side ZIP Codes, ridership remained low over 5 years

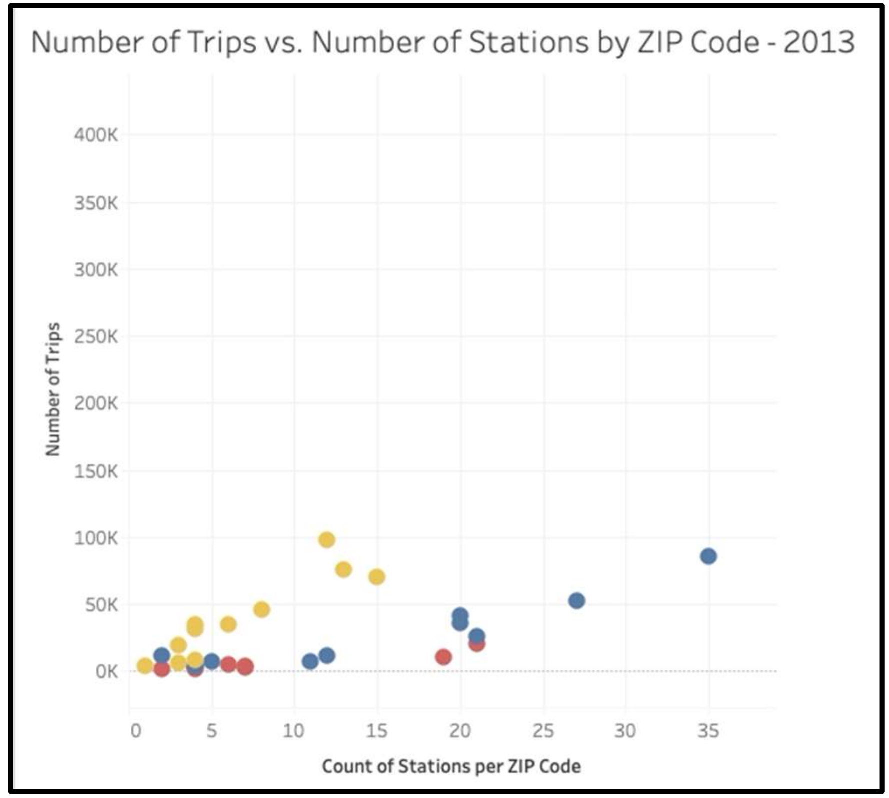


Figure 2 (video)

Increasing the number of stations alone did not increase ridership

Price

Hypothesis: Riders in low-income areas may not be able to afford $99 annual subscription, therefore decreasing the price would increase ridership

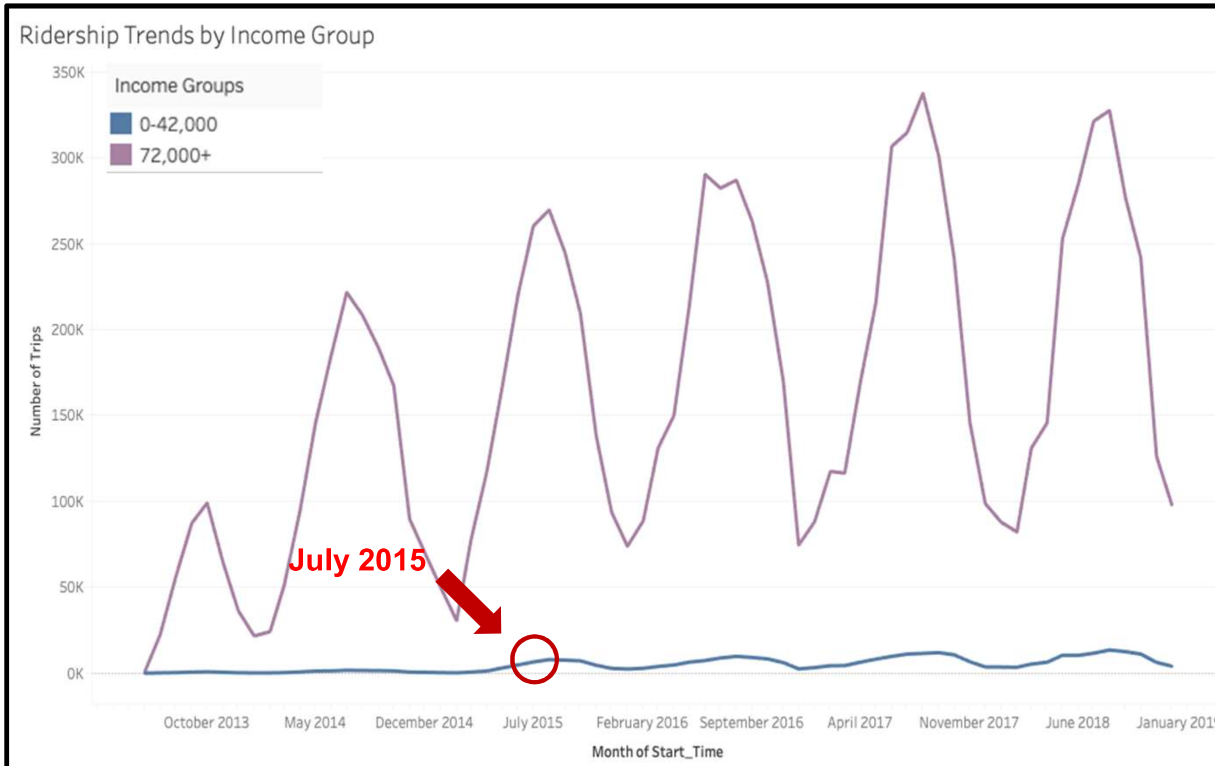
Divvy for Everyone (D4E) launched in July 2015

- Initiative started by Divvy to encourage usage by low income riders
- $5 subscription for first year, gradual increase in price until Year 4, subscription capped at $75

Figure 3 shows the trend of subscriber rides in low income areas and high income areas

Small increase in low-income rides after D4E launched, but all income groups saw slight growth from 2014 to 2015

Program didn't make large impact closing the gap between income groups



Lowering the price for a subscription did not result in any significant increase in rides in low-income areas

UNIVERSE DENVER 2019

Time Limit

Hypothesis: Low-Income areas are less concentrated than the downtown area and rides may take longer than the allotted time resulting in overage charges

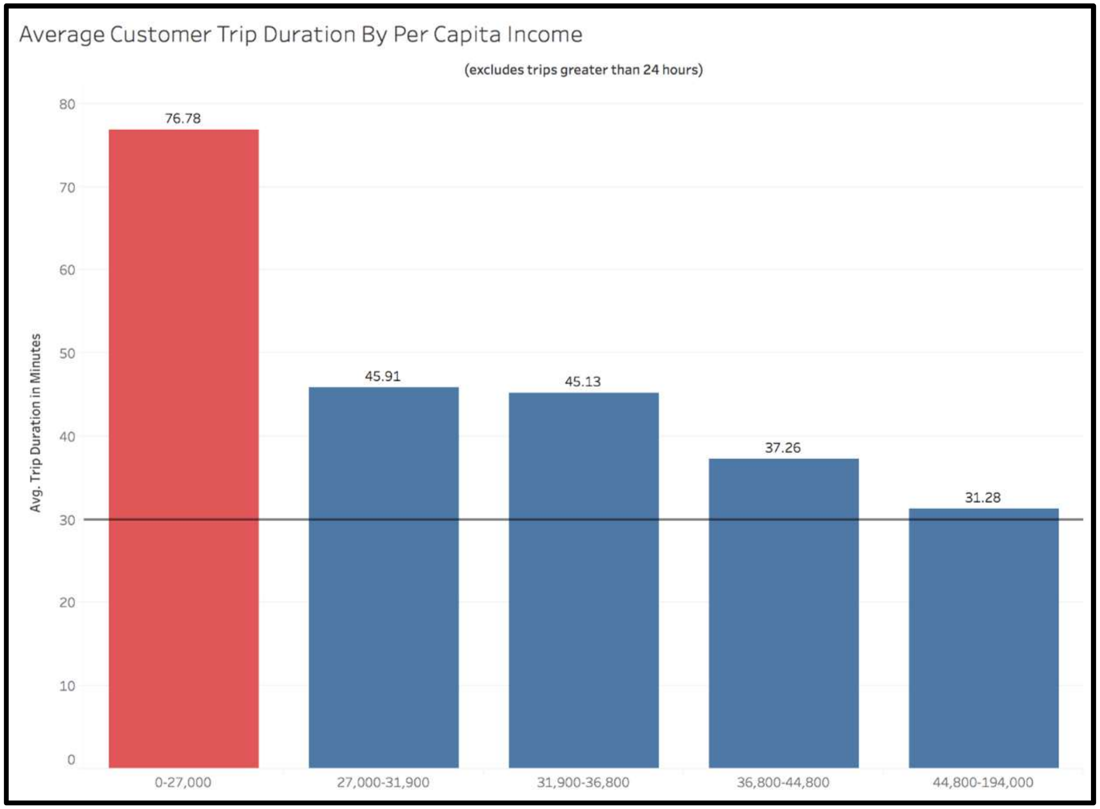
Time Limits

- Customer (1-time rider): 30 minutes
- Subscriber (annual pass): 45 minutes

Customer trips cost $3 per 30 minute ride, with an addition $3 charge for every 30 minutes over limit

Figure 4 shows that trips occurring in low income areas last about 77 minutes, more than double the time limit, making the average trip costing $9

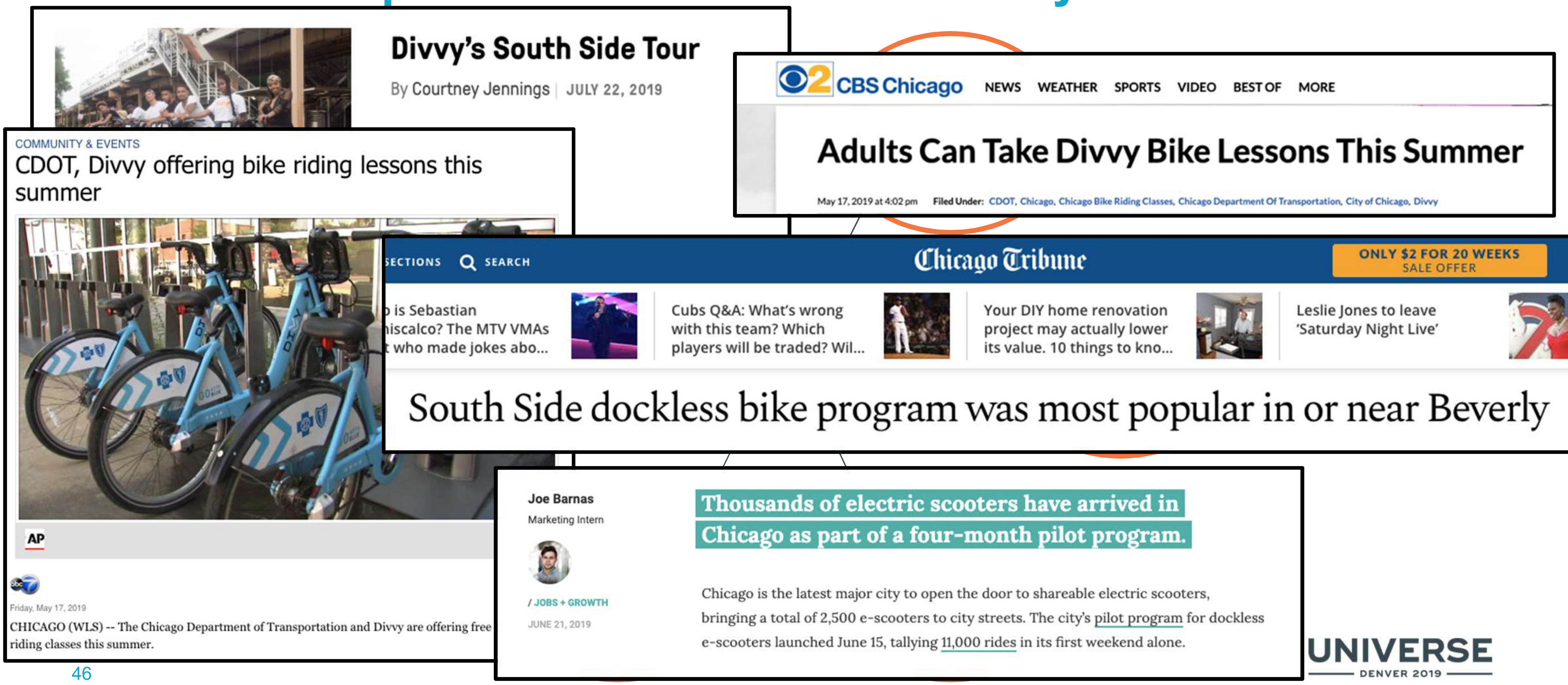
*train or bus ticket costs $2.50 - $3.00



Customer trips in low-income areas are more than double the time limit of 30 minutes

UNIVERSE DENVER 2019

Identifying these barriers in isolation and only addressing some of them has not proven effective, instead a multifaceted approach is required that considers a variety of factors



UNIVERSE
DENVER 2019

Thank you.

A5- Southern New Hampshire University USA

Food Insecurity and Waste

Presenter - Mitchell Beckner

Team members – Alexander Hilton, Jonathan Squire

Faculty Advisor – Matt Keane

October 2019

UNIVERSE

DENVER 2019

teradata.

snhu

Our Vision

To create a strategy that builds a better, properly fed world, reducing waste and bringing meals to those in need

US Food Supply

- Current world population is 7.7 billion
- Global food production is enough to feed 10 billion
- Current US population is 327.2 million
- The US is the world's 3rd largest food producer

Is there enough?

UNIVERSE
DENVER 2019

Food Insecurity

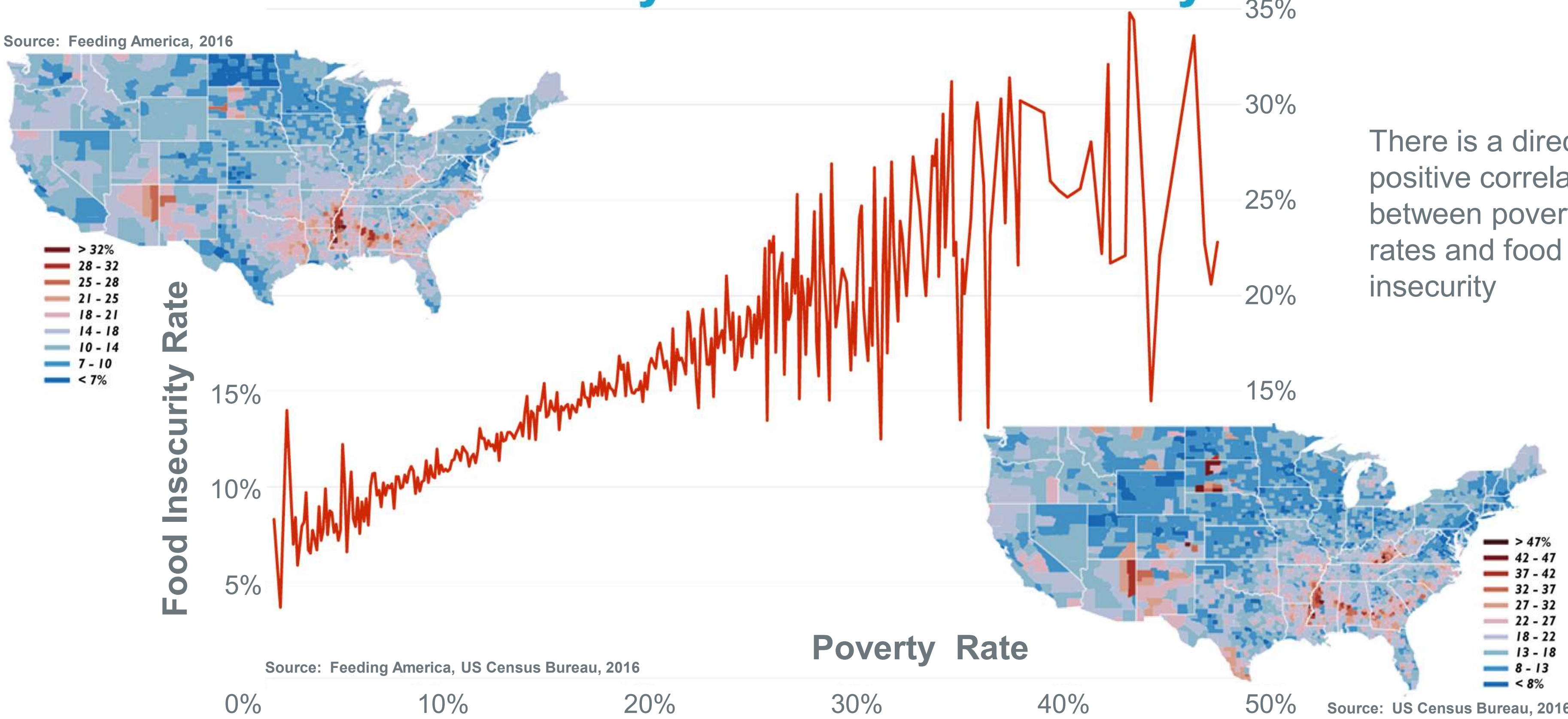
What is it?

The state of being without reliable access to a sufficient quantity of affordable, nutritious food

Is it a problem?

1 in 8 Americans are affected;
in some communities, 1 in 6

The Influence of Poverty on Food Insecurity

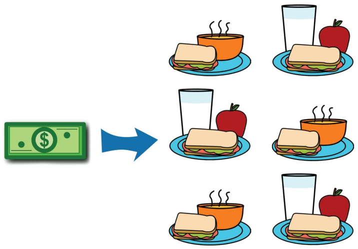


There is a direct, positive correlation between poverty rates and food insecurity

UNIVERSE DENVER 2019

Reducing the Costs

Donations of Money



"A dollar can go a long way. For every dollar donated, we provide six meals."

- Lora D. – The Foodbank

Donations of Time

"Volunteering saves $2.5 million annually. [That] money could be spent on groceries."

- Maureen M – Second Harvest

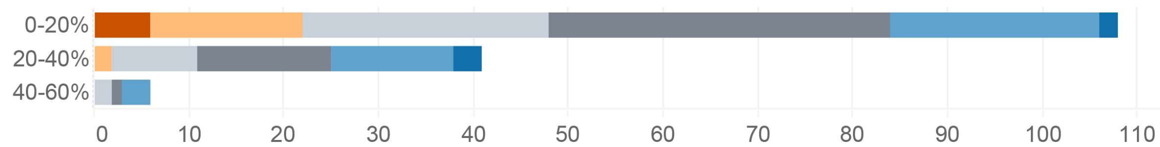
Is Food Waste a Problem?

Most commonly wasted foods

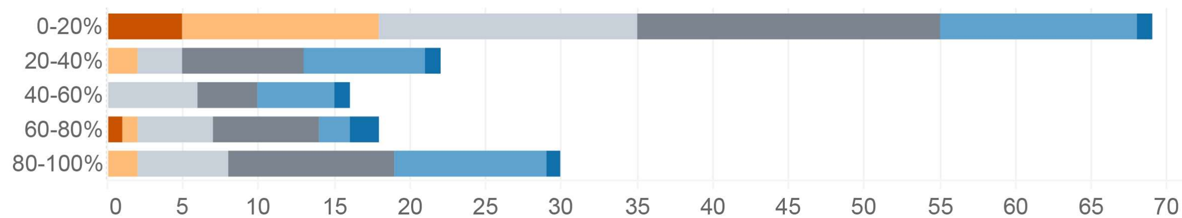
Source: IT476 Food Waste Survey

Age	Respondants
18-24	6
25-34	39
35-44	51
45-54	37
55-64	18
65-74	6
Total	157

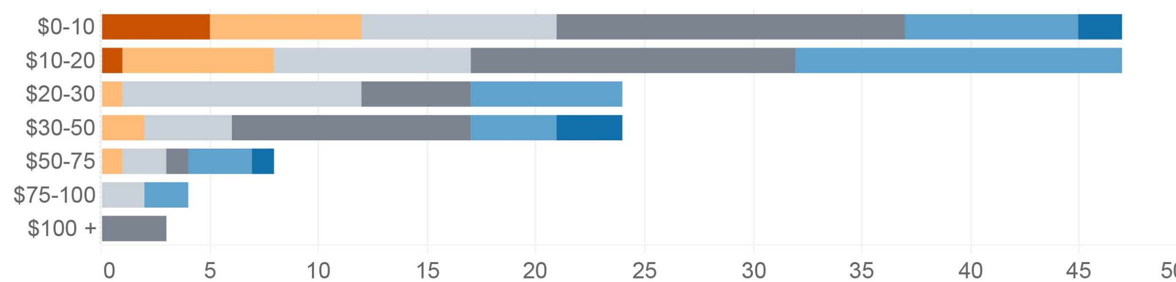
How much do we waste?



How much could we prevent?

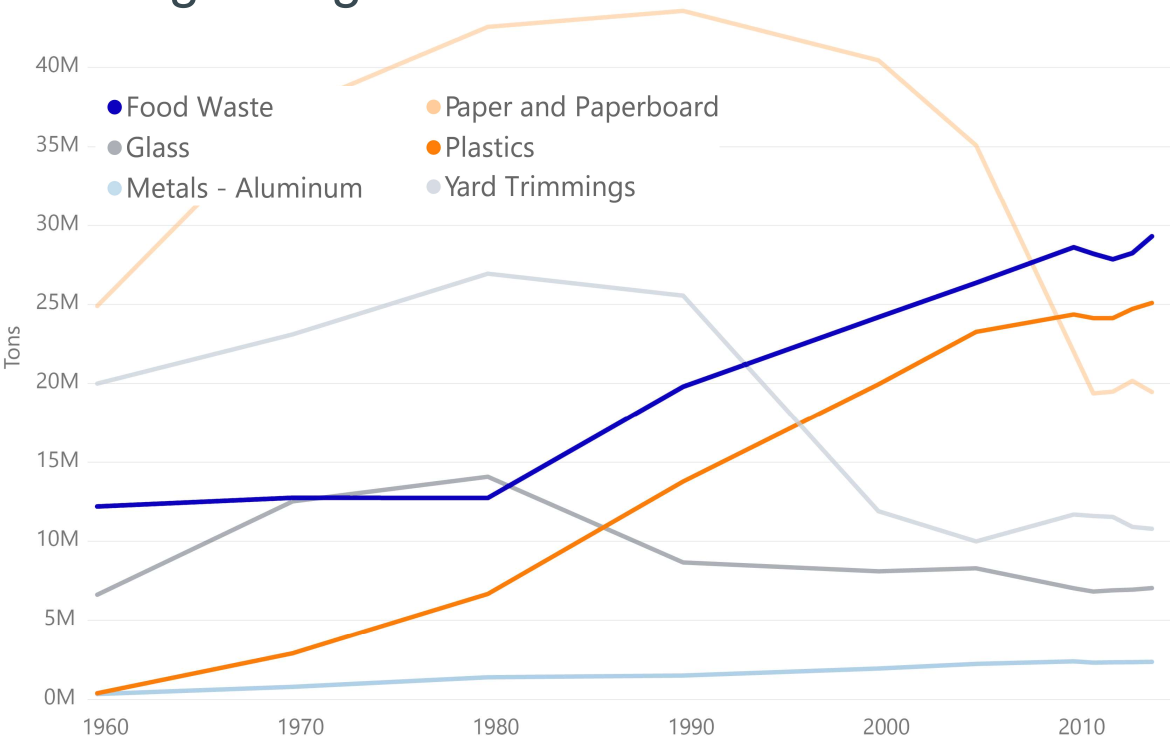


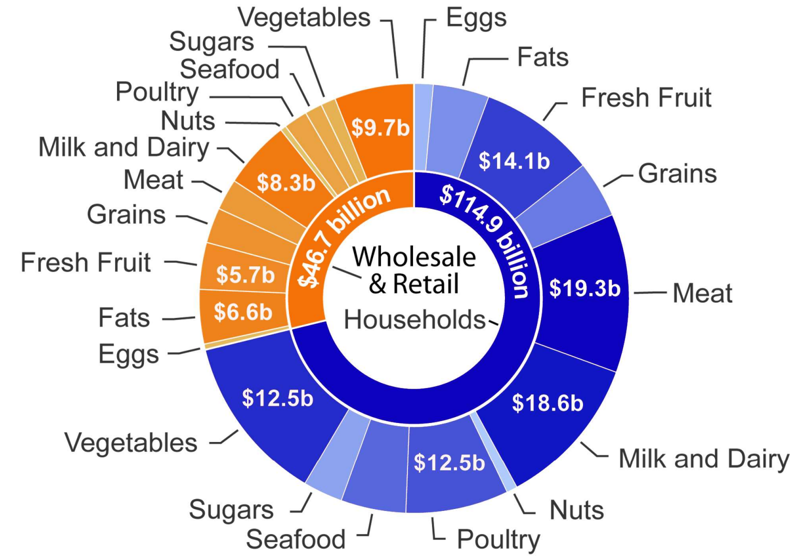
How much could we save?



US Food Waste Figures

A growing trend





Source: Buzby et al., 2014

Americans waste 150,000 tons of food per day

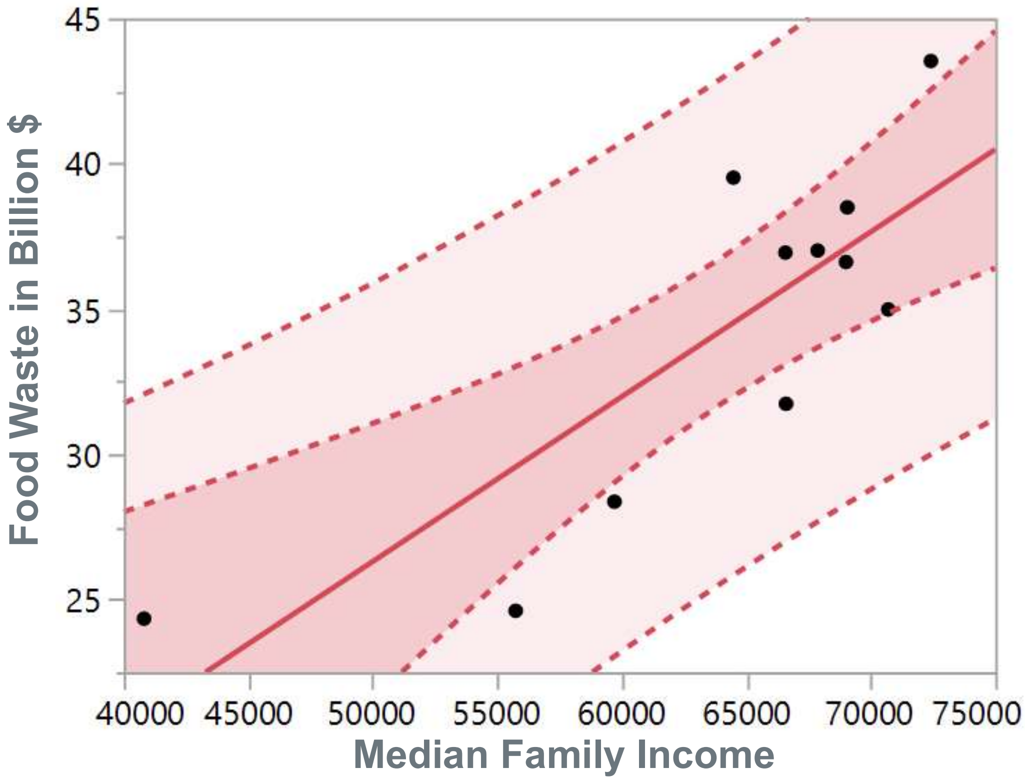
Food waste is estimated at 30-40% of the total food supply

UNIVERSE
DENVER 2019

The Influence of Income on Food Waste

As median family income increases, so does the amount of food wasted

Individuals with healthier diets are often the most wasteful



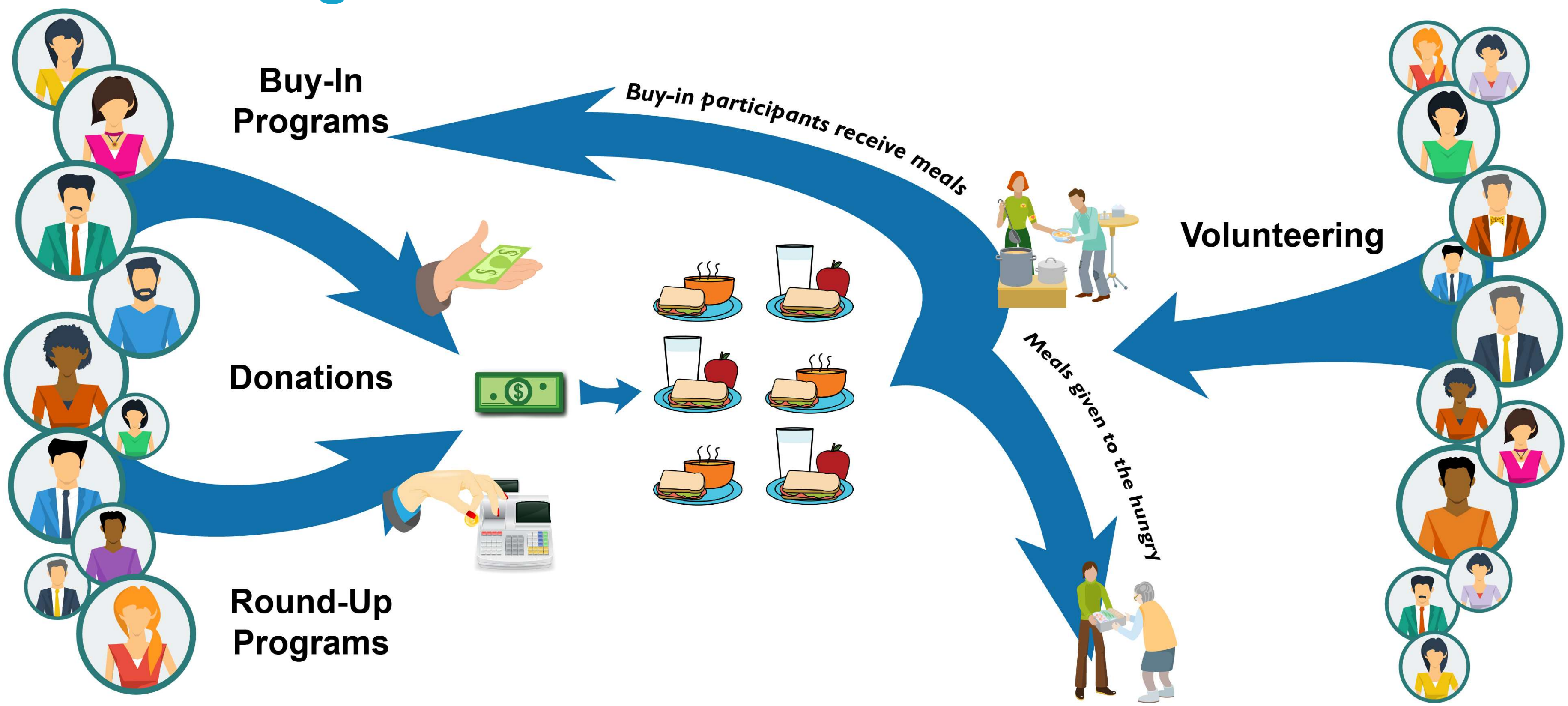
Sources: US Census Bureau, Food and Agriculture Organization of the United Nations

Taking Action!

Actions everyone can take

- Support standardization of regulations
- Volunteer your time
- Shop wisely
- Donate what and when you can

Our Challenge to You!



Thank you.

- Lora Davenport, Advocacy & Programs Manager at The Foodbank Inc. in Dayton, OH

  Lora provided insightful firsthand information on the challenges of food insecurity and the various ways that her organization helps those in need in the Dayton area.

- Maureen Mikel, Development Manager at Second Harvest Food Bank of Central Florida in Orlando, FL

  Maureen advised of ways that Second Harvest helps those in need in the community, including programs specifically aimed at ensuring children of families in need receive meals both in school and on the weekend between school.

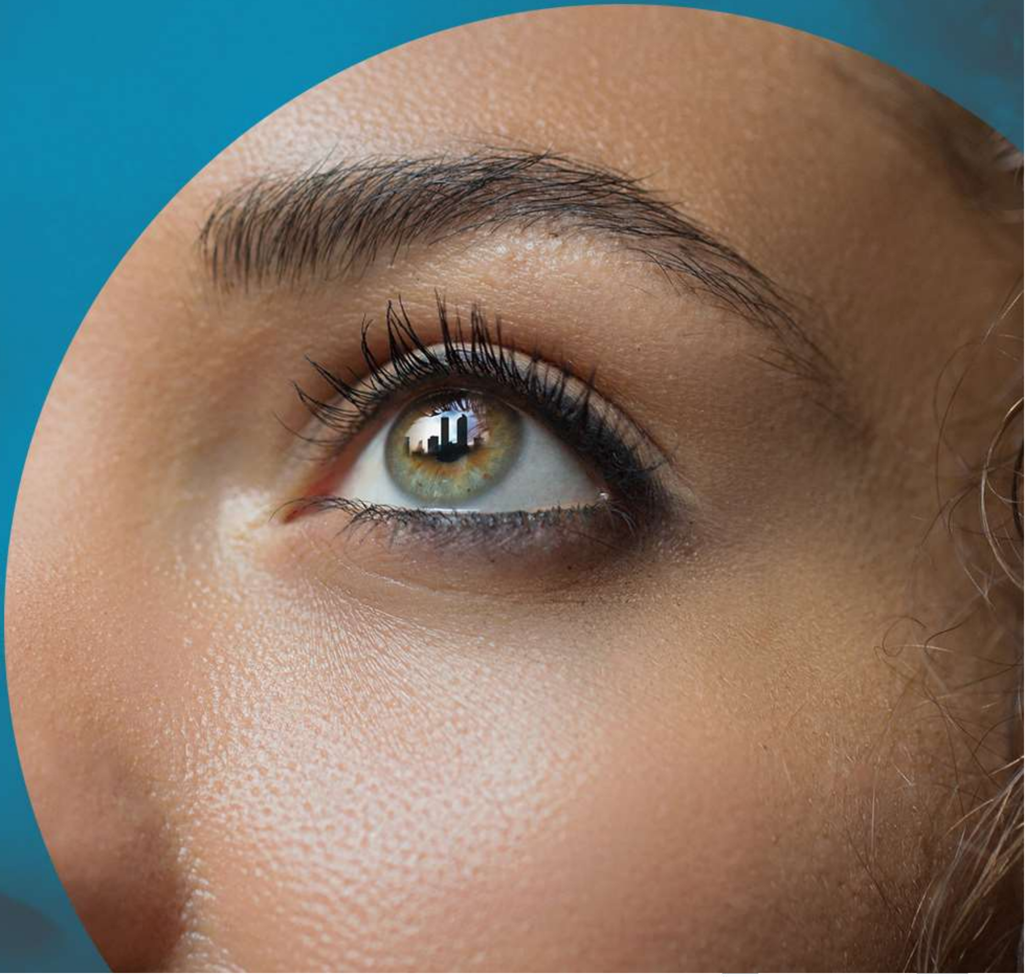
- The Foodbank Inc: thefoodbankdayton.org
- Second Harvest Food Bank of Central Florida: www.feedhopenow.org
- To the 169 friends, family, and fellow students from SNHU who filled out the survey

A6- University of Minnesota Duluth USA

Impact of Health Disparities on the Economy and Communities in Duluth, Minnesota

Noah Lahr
Patrick O’Brien
October, 2019

UNIVERSE
DENVER 2019
teradata.



Agenda

- Project Significance
- Defining Health Status
- Analytics Approach – Ideation, Data and Analysis
- Cluster Analysis
- Analytic Insights
- Moving Forward

Project Significance

“Minnesota is one of the healthiest states in the country, but it has some of the worst health disparities”. The Cost: $2.26 Billion Annually.

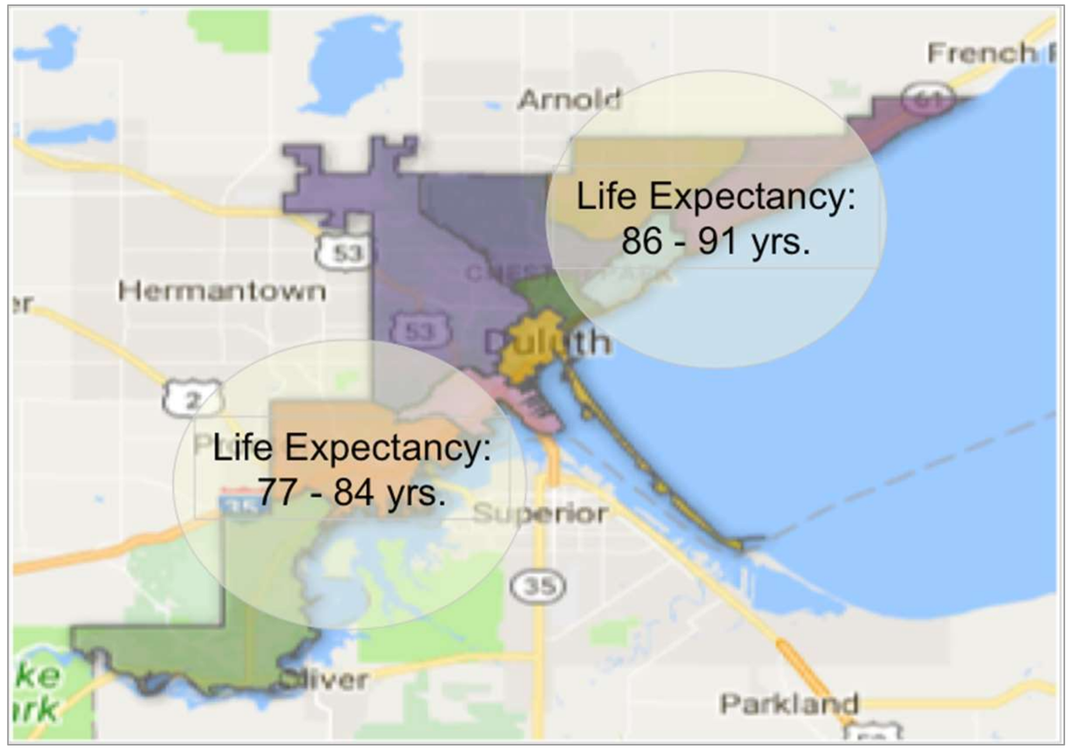
– BlueCross BlueShield MN
2018 Technical Report

UNIVERSE
DENVER 2019

Project Significance

In Duluth, health disparities have persisted over the last decade. Where you live in Duluth can lessen your life expectancy by as much as 14 years.*

*2017 *Health Status Report Update to the Community. Life expectancies determined by confidence intervals of geographically defined neighborhoods.*



Defining Health Status

Social Determinants of Health:

- Income
- Poverty
- Employment
- Education

- Preventable Diseases
- Poverty
- Cost Burden Households
- Race/Ethnicity

Figure from https://www.bluecrossmn.com/healthy/public/portalcomponents/PublicContentServlet?contentId=P11GA_16928228

Analytics Approach

Cluster analysis on census data factors of social determinants of health to remove artificial construct of geographically connected neighborhoods.

Use clusters to model variance of health disparities across the city of Duluth and provide insight into resource/access disparity.

Use State level analysis methodology of economic loss to approximate the impact on an annual basis to Duluth.

Data

- **2005 – 2017 Mortality Data** (Minnesota DPH)

Data included addresses of decedents, causes of death (as up to 20 ICD10 codes), and extensive demographic information (gender, age, race, occupation, industry)

- **U.S. Census Bureau ACS 5-year estimates (2013-2017)**

Population demographics, social, and housing/household characteristics for all census tracts in Duluth

- **St. Louis County Geospatial Data**

Housing prices, residential density, geography of parks, community use land/buildings, grocery, gas, convenience, and liquor stores.

- **Nutritional Assistance Data (St. Louis County DPH)**

Information on locations where Supplemental Nutrition Assistance Program benefits could be used (EBT stores).

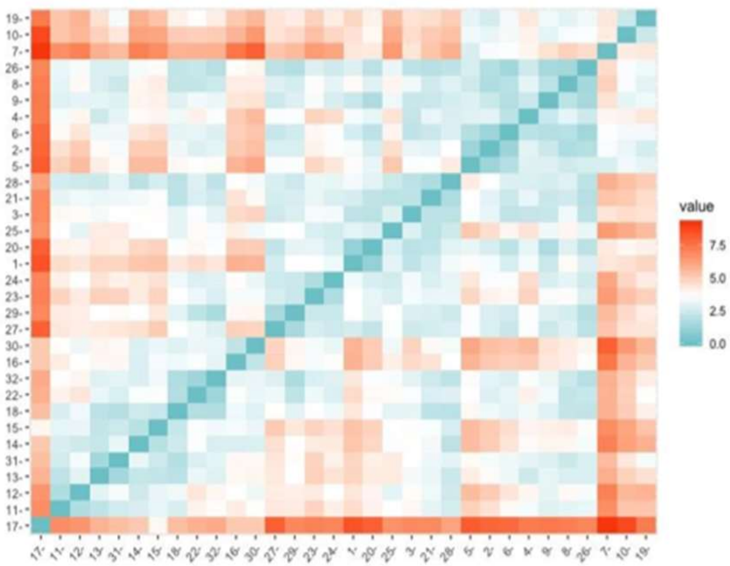
Data Cleaning

- Geocode decedents to census tracts to identify Duluth residents
- Coded occupation/wage data using NAICS Standards
- Coded cause of death
- Normalized Social Determinant Factors

Analytics

- Probability modeling - life expectancies/economic impact
- Principal component analysis, regression, and correlation were used to validate variable selection used to perform k-Means clustering on census tract data.

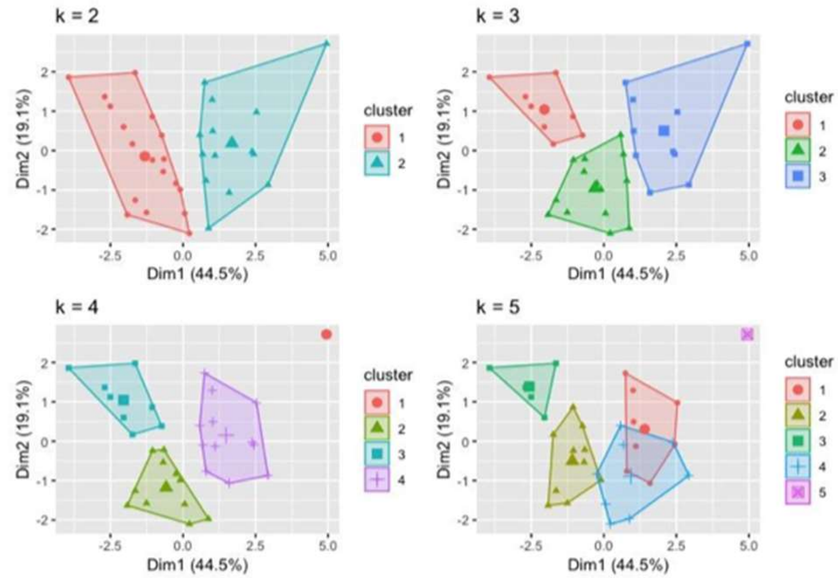
Cluster Analysis



Distance Matrix:

Red = large difference between census tracts
Teal = large similarity between tracts

Result: Clustering is useful.

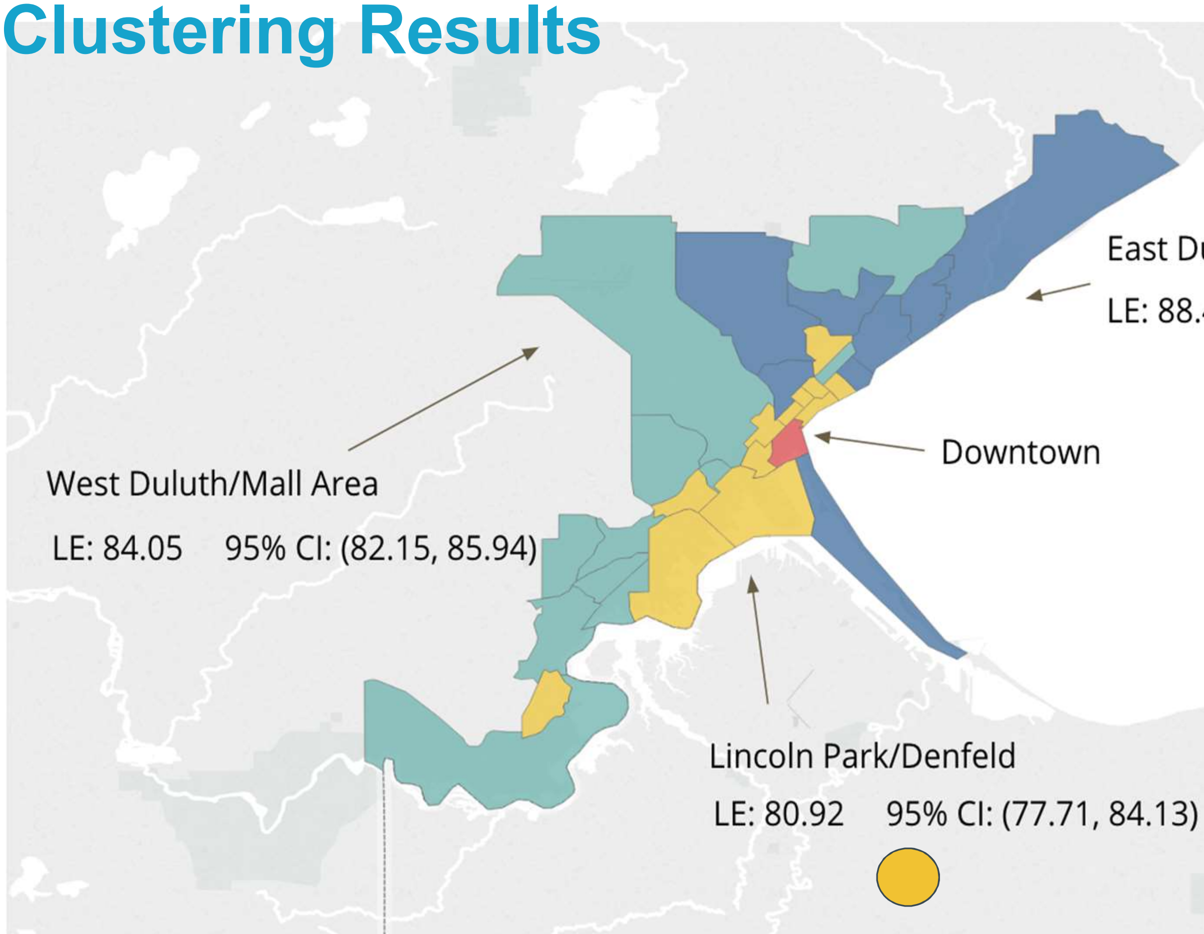


Optimal: 4 clusters

Principal component to determine the optimal number of clusters.

UNIVERSE DENVER 2019

Clustering Results



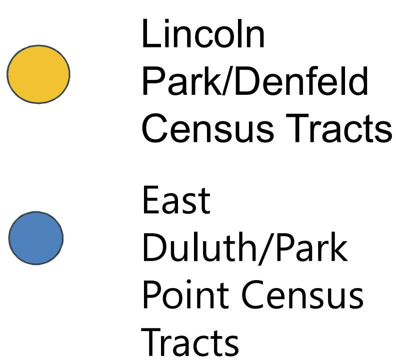
Above average Income	$72,334
Above average House Prices	$204,150
Above average Education	97.24%
Below average Poverty	13.43%
Below average Unemployment	2.48%

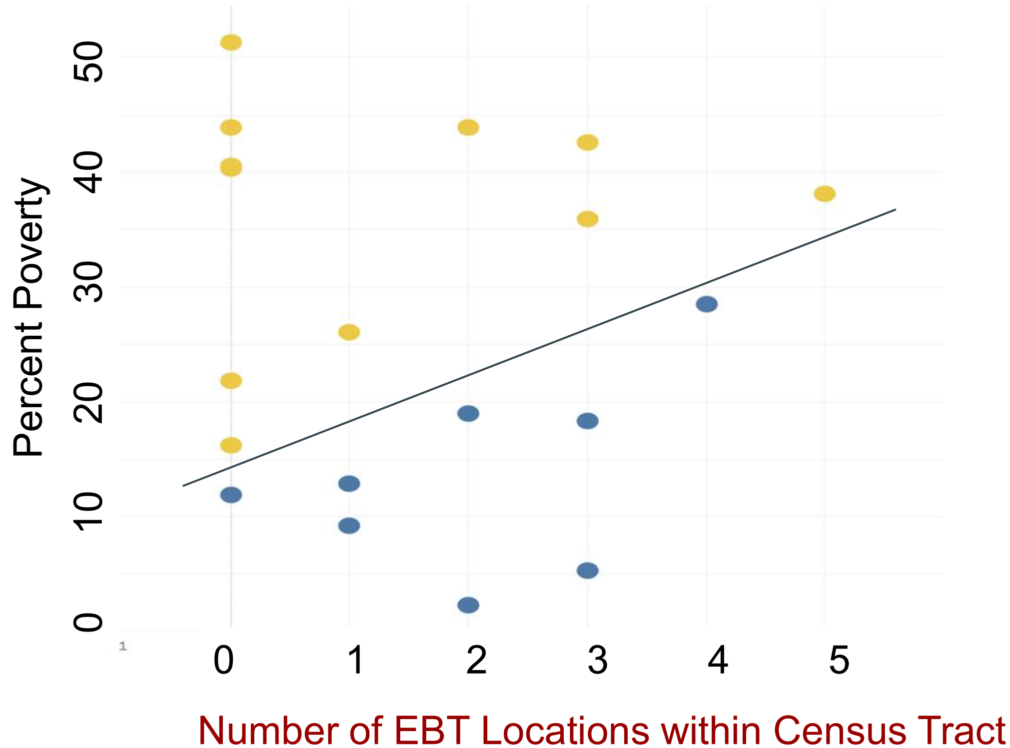
Below average Income	$31,999
Below average House Prices	$111,114
Below average Education	91.45%
Above average Poverty	34.73%
Above average Unemployment	4.67%

UNIVERSE DENVER 2019

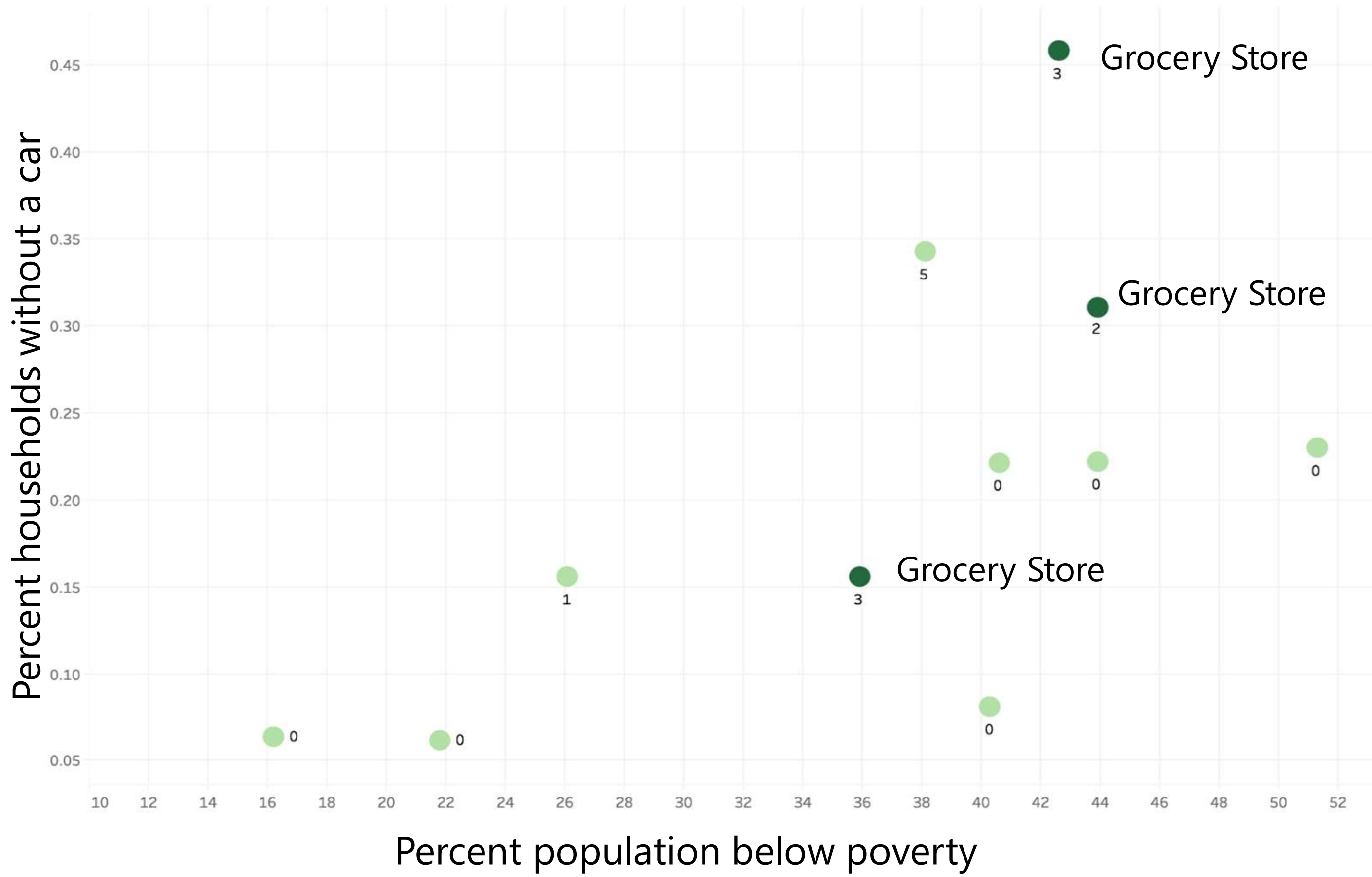
Resource Analysis

Food Scarcity:





Resource Analysis



UNIVERSE
DENVER 2019

Economic Analysis

2017 Economic Loss Analysis*	Average Age of Death	Average Annual Income	Gross Annual Economic Loss
East Duluth/ Park Point	67	$58,985	$1.8 Million
Lincoln Park/ Denfeld	59	$57,445	$2.8 Million
City of Duluth	62	$51,317	**$8.1 Million**

*Age: 25-74 (working aged defined by Bureau Labor & Statistics)
Cause of Death = Heart Disease, Cancer, Diabetes, Alcohol Abuse, Suicide

Moving Forward

Medical Care

- ✓ Optimize location of preventive care facilities to increase access
- ✓ Improve access to/routes of public transportation

Food Access

- ✓ Incentivize full grocery retailers to open in areas of high poverty
- ✓ Improve access to/routes of public transportation

Education

- ✓ Assess impact of community based continuing education for job training/career transition

Housing

- ✓ Address lack of affordable housing
- ✓ Develop plan to improve built environments of cost burdened households

Thank you.

UNIVERSE
DENVER 2019

ANALYTICS CHALLENGE WRAP UP

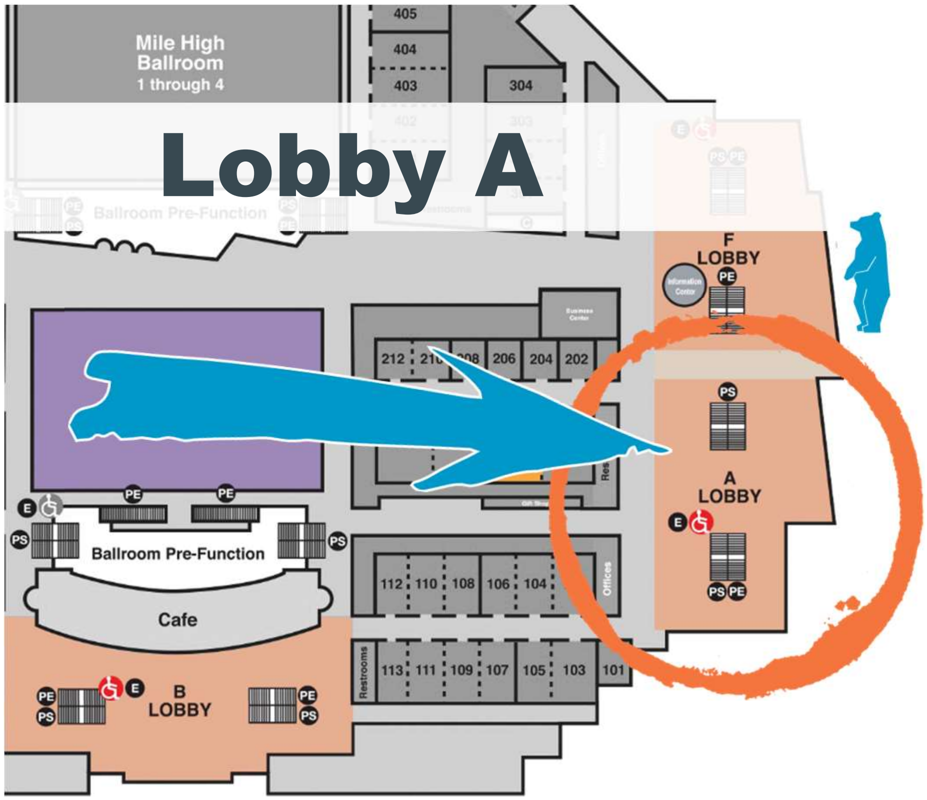
UNIVERSE

DENVER 2019

teradata.

Student Poster Presentations

Meet All Finalists and Honorable Mention teams!!



Sunday
Oct 20

6:30 – 8:00pm

SUNDAY

MONDAY

Monday
Oct 21

8 – 8:45am

11:00am – 1:45pm

UNIVERSE
DENVER 2019

Award Categories

Data Challenge

• **People's Choice - Best Presentation**	**Attendees vote**
• Overall Winner	Academic vote
• Best Value to Hire Heroes USA	Hire Heroes USA vote

Analytics Challenge

• **People's Choice - Best Presentation**	**Attendees vote**
• **Best Use of Analytics and Visualization Tools**	**Attendees vote**
• Overall Winner	Academic vote

Teradata Technology Award

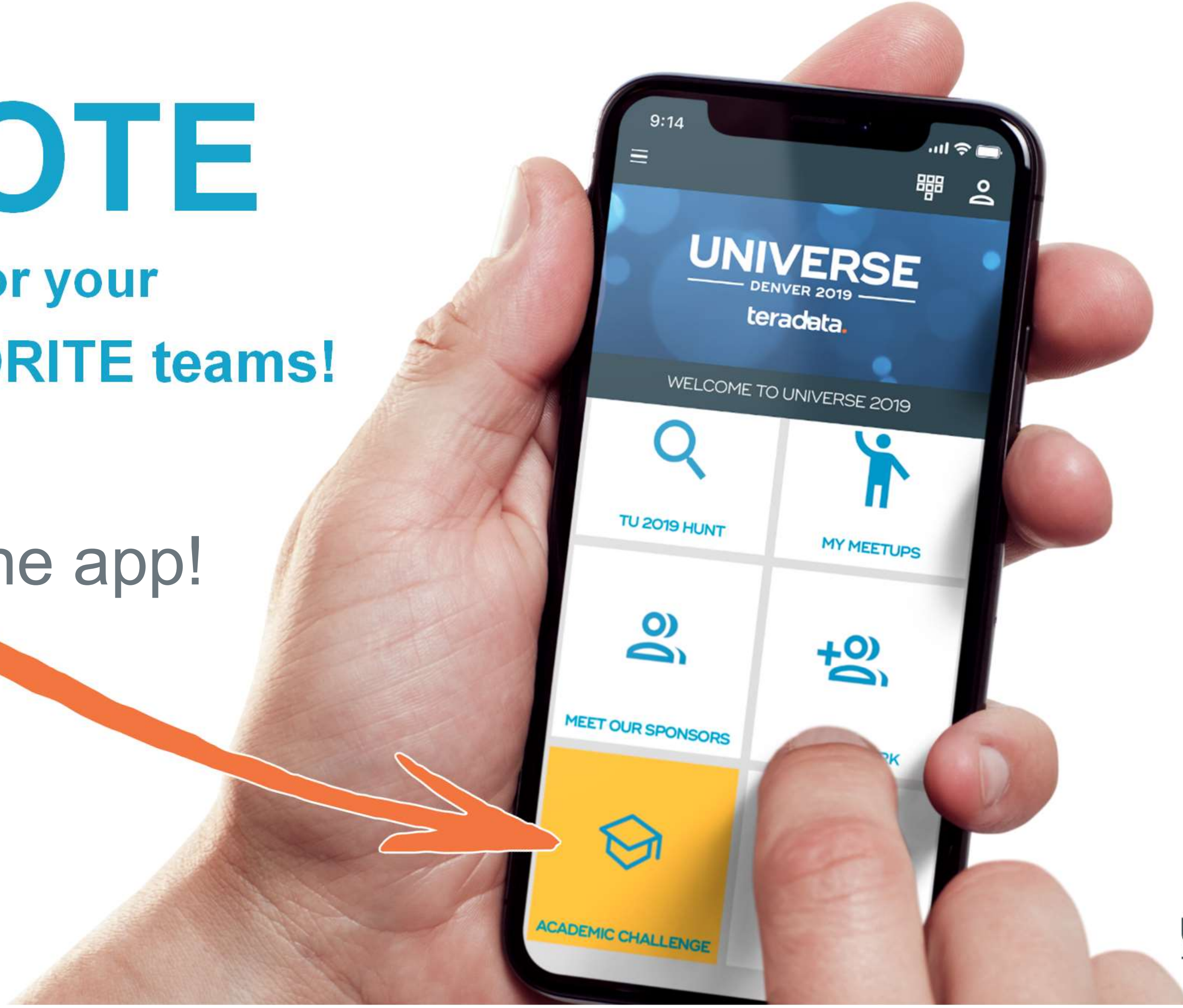
• Best Use of Teradata Technology	Academic vote

VOTE

for your

FAVORITE teams!

It's in the app!



Voting **opens** on **Sunday** at 3pm

Voting closes on **Monday** at 2pm

UNIVERSE
DENVER 2019

Student Celebration Event

Monday, October 21, 6:30-8:30pm

All attendees invited!

UNIVERSE
DENVER 2019

Thank you.

UNIVERSE
DENVER 2019