

---

# Rates of Diabetes in the United States

Poseidon Team



# The Poseidon Team



**Katejoe Akabogu**

---

Business Intelligence  
Analyst, Chevron

Fun fact: loves to juice and  
bake; zumba dance



**Chris Chaney**

---

Data Analyst

Fun fact: self proclaimed  
Pokémon Master



**Liat Sayfan**

---

Project Manager, Applied  
Survey Research

Fun fact: teaches spin  
classes; blue belt in BJJ



**Farrukh Sultani**

---

Managing Assistant- Big 5  
Corps

Fun fact: loves smiling just like  
the meaning of my name.

# Outline

- Overview of the project
- Research questions
- Process
- Understanding the datasets
- Descriptive analytics
- Predictive analytics
- Summary of findings

---

# Overview

Diabetes is a chronic disease that affects how the body turns food into energy. There are three main types of diabetes - type 1, 2, and gestational diabetes (diabetes while pregnant). According to CDC, more than 133 million in the United States live with diabetes, and 96 million out of 133 are US adults.

Based on studies conducted in the United States so far, there isn't a specific cure for diabetes yet, but weight management, good nutrition, and being active can improve the health outcome of a diabetic patient.

While previous studies focused on data at the individual level, the focus of our study was to examine the relationship between physical activity, food nutrition, and obesity on diabetes prognosis at the state level.



---

# Research Questions

1

What is the rate of diabetes in the United States?

2

Are there any differences in rates of diabetes between the four major regions of the United States?

3

How strong is the relationship between nutrition, physical activity, obesity and diabetes at the state level?

4

Can we predict diabetes rates at the state level based on nutrition rate, physical activity rates, and obesity rates?

# Process



## Data Wrangling & Transformation

- Clean diabetes rate dataset.
- Clean nutrition, physical activity and dataset.
- Merged the two clean datasets.

## Descriptive Analytics

- Describe the rates of diabetes in United States.
- Compare rates among United States regions.



## Predictive Analytics

- Inspect relationship between nutrition, physical activity, obesity, and diabetes.
- Predict diabetes rates based on nutrition, physical activity, obesity and diabetes.



# Understanding the Datasets



## Challenges

When researching the topic, we found multiple data sources but few captured the data we wanted to use.

We found two datasets that had the variables of interest, but they had different structures, and we needed to transform them to be a similar structure in order to merge them



# Dataset 1

## 500 Cities: Adult Diabetes<sub>(csv)</sub>

Data was provided by the Centers for Disease Control and Prevention (CDC), Division of Population Health, Epidemiology and Surveillance Branch. This project explores diabetes rates at the city level and includes sample size and population

### Variables & Observations:

This dataset displays 24 columns, 29,006 rows in table. This dataset measures 50 states and 2 other US owned territories.

Diagnosed diabetes among adults aged >=18...

ties: Local Data for Better Health, 2019 release

provided by the Centers for Disease Control and Prevention (CDC), Division ▶

te... :	State... :	CityN... :	Geograph... :	Data... :	Categ... :	Uniqu... :	Measure :
	California	Inglewood	Census Tract	BRFSS	Health Ou...	0636546...	Diagnosed dia...
	Alabama	Huntsville	Census Tract	BRFSS	Health Ou...	0137000...	Diagnosed dia...
	Alaska	Anchorage	Census Tract	BRFSS	Health Ou...	0203000...	Diagnosed dia...
	Arizona	Tempe	Census Tract	BRFSS	Health Ou...	0473000...	Diagnosed dia...
	Arizona	Yuma	City	BRFSS	Health Ou...	0485540	Diagnosed dia...
	California	Antioch	Census Tract	BRFSS	Health Ou...	0602252...	Diagnosed dia...
	California	Burbank	Census Tract	BRFSS	Health Ou...	0608954...	Diagnosed dia...
	California	Folsom	City	BRFSS	Health Ou...	0624638	Diagnosed dia...
	California	Moreno V...	City	BRFSS	Health Ou...	0649270	Diagnosed dia...
	California	Palmdale	Census Tract	BRFSS	Health Ou...	0655156...	Diagnosed dia...
	California	Riverside	Census Tract	BRFSS	Health Ou...	0662000...	Diagnosed dia...
	California	Roseville	Census Tract	BRFSS	Health Ou...	0662938...	Diagnosed dia...
	California	San Bruno	City	BRFSS	Health Ou...	0667000	Diagnosed dia...



# Dataset 2

## Nutrition, Physical Activity, and Obesity (nPAO.csv)

2011 - 2015 data by state from the CDC Behavioral Risk Factor Surveillance System (BRFSS). Covers questions on nutrition, physical activity, and obesity.

### Variables & Observations:

This dataset displays 33 columns, 48,772 rows in table.  
This dataset measures 50 states and 4 other US owned territories.

### Original

```
nPAO_original.head()
```

	YearStart	YearEnd	LocationAbbr	LocationDesc	Datasource	Class	Topic	Question	Data_Value_Unit	Data_Value_Type	...	GeoLocation	Cl
0	2011	2011	AL	Alabama	Behavioral Risk Factor Surveillance System	Obesity / Weight Status	Obesity / Weight Status	Percent of adults aged 18 years and older who ...		NaN	Value	(32.8405711220048, -86.63186076199969)	
1	2011	2011	AL	Alabama	Behavioral Risk Factor Surveillance System	Obesity / Weight Status	Obesity / Weight Status	Percent of adults aged 18 years and older who ...		NaN	Value	(32.8405711220048, -86.63186076199969)	
2	2011	2011	AL	Alabama	Behavioral Risk Factor Surveillance System	Obesity / Weight Status	Obesity / Weight Status	Percent of adults aged 18 years and older who ...		NaN	Value	(32.8405711220048, -86.63186076199969)	
3	2011	2011	AL	Alabama	Behavioral Risk Factor Surveillance System	Obesity / Weight Status	Obesity / Weight Status	Percent of adults aged 18 years and older who ...		NaN	Value	(32.8405711220048, -86.63186076199969)	
4	2011	2011	AL	Alabama	Behavioral Risk Factor Surveillance System	Obesity / Weight Status	Obesity / Weight Status	Percent of adults aged 18 years and older who ...		NaN	Value	(32.8405711220048, -86.63186076199969)	



## Dataset 2

### Nutrition, Physical Activity, and Obesity (nPAO.csv)

#### Reformatting

LocationDesc	GeoLocation	Class	Question	Value	XSample	Sample_Size
Alabama	(32.84057112200048, -86.63186076199969)	Fruits and Vegetables	Percent of adults who report consuming fruit less than one time daily	5311147.2	116465.0	
			Percent of adults who report consuming vegetables less than one time daily	3051087.1	119184.0	
		Obesity / Weight Status	Percent of adults aged 18 years and older who have an overweight classification	7766780.1	225761.0	
			Percent of adults aged 18 years and older who have obesity	7546136.6	225798.0	
		Physical Activity	Percent of adults who achieve at least 150 minutes a week of moderate-intensity aerobic physical activity or 75 minutes a week of vigorous-intensity aerobic activity (or an equivalent combination)	5286812.9	120273.0	
			Percent of adults who achieve at least 150 minutes a week of moderate-intensity aerobic physical activity or 75 minutes a week of vigorous-intensity aerobic activity (or an equivalent combination)	5474756.7	101182.0	
... Wyoming	(43.23554134300048, -108.10983035299967)	Physical Activity	Percent of adults who achieve at least 150 minutes a week of moderate-intensity aerobic physical activity or 75 minutes a week of vigorous-intensity aerobic activity (or an equivalent combination)	2128092.1	100136.0	
			Percent of adults who achieve at least 150 minutes a week of moderate-intensity aerobic physical activity or 75 minutes a week of vigorous-intensity aerobic physical activity and engage in muscle-strengthening activities on 2 or more days a week	3533692.7	100255.0	
			Percent of adults who achieve at least 300 minutes a week of moderate-intensity aerobic physical activity or 150 minutes a week of vigorous-intensity aerobic physical activity and engage in muscle-strengthening activities on 2 or more days a week	3533692.7	100255.0	



## Dataset 2

Nutrition, Physical Activity,  
and Obesity ([nPAO.csv](#))

Final

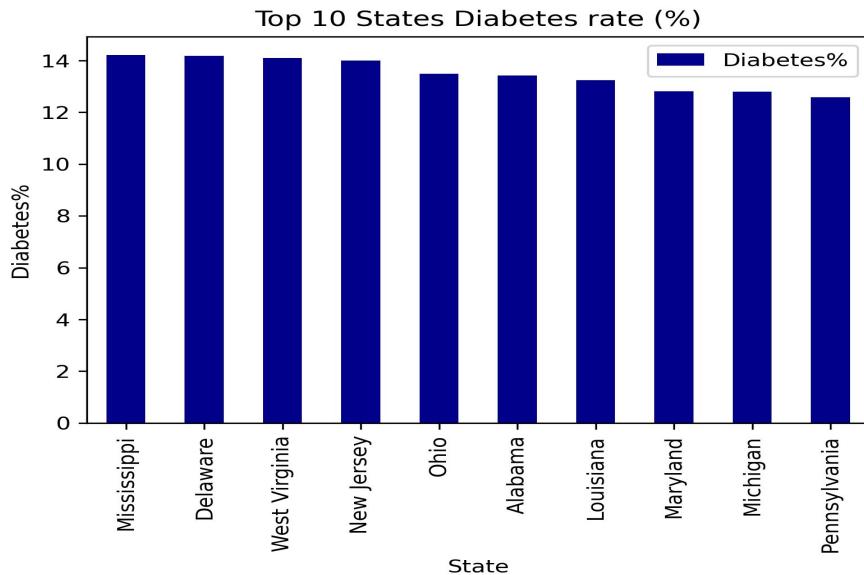
	State	Long & Lat	Obesity %	Overweight %	In-Active %	Muscle Strengthening %	Low Aerobics %	High Aerobics %	Strength & Conditioning %	Low Veggies Consumption %	Low Fruit Consumption %
0	Alabama	(32.84057112200048, -86.63186076199969)	33.419856	34.402665	30.351719	24.379398	43.956773	25.801610	15.482848	25.599805	45.602947
1	Alaska	(64.84507995700051, -147.72205903599973)	28.337146	37.627901	20.651719	34.140096	57.327324	37.255386	24.586290	18.776312	39.087520
2	Arizona	(34.865970280000454, -111.76381127699972)	27.033482	36.211155	22.765078	31.002902	53.824667	34.435845	22.613209	20.743749	38.150600
3	Arkansas	(34.74865012400045, -92.27449074299966)	33.932269	34.859221	32.900491	22.912310	43.908805	27.368902	14.846760	27.518560	48.642681
4	California	(37.63864012300047, -120.9999953799971)	24.056118	35.882303	19.668067	32.605964	58.273079	36.688512	23.774697	16.757566	31.444280



# Descriptive Analytics



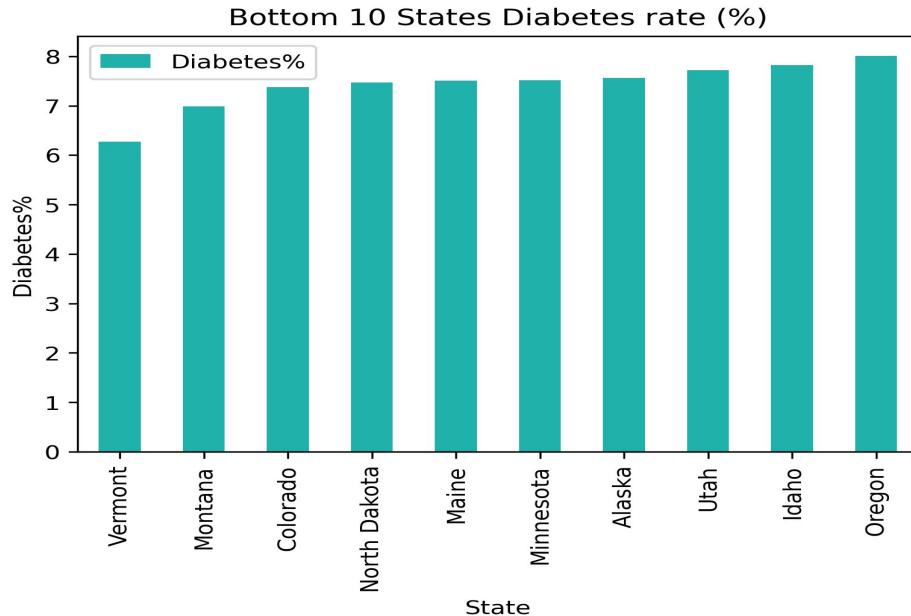
# Rate of Diabetes in the US



Most of the south region states with few northeast & midwest region states indicated high diabetes rate.

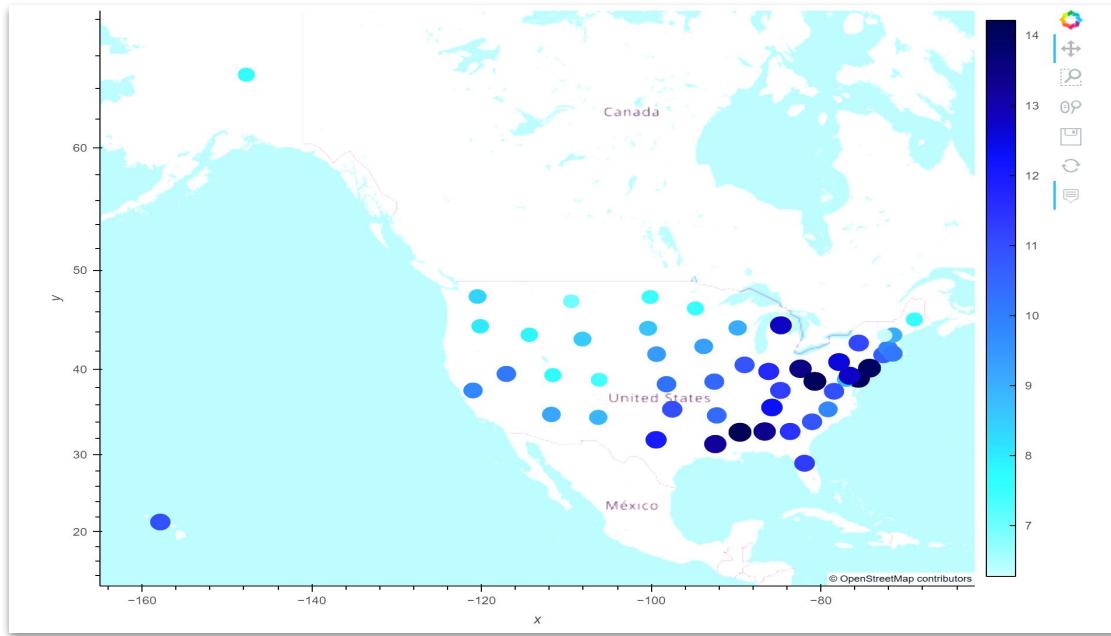


# Rate of Diabetes in the US

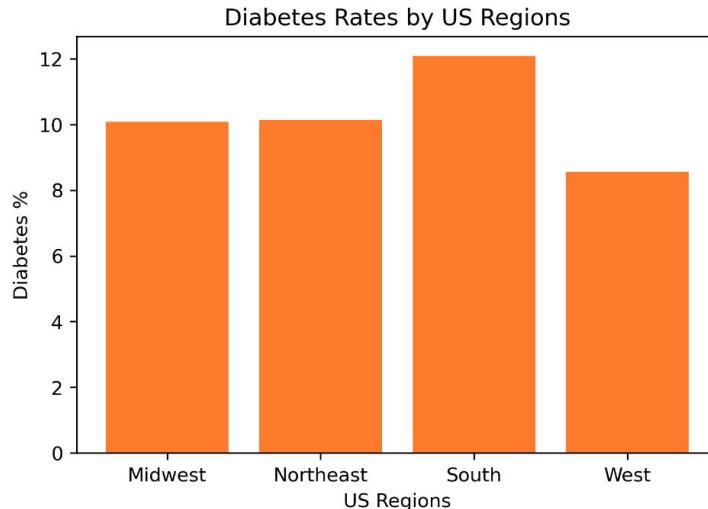


The West region states (combination of West and Midwest) showed low diabetes rates, with the exception of Vermont and Maine state that are in the Northeast region.

# Map of U.S Diabetes Rate



# Compare rates among US regions



There is a significant difference in the rates of diabetes between the different US regions.

According to our graph, the west region has the lowest diabetes rates, and the south has the highest rates of diabetes.

F\_onewayResult(statistic=10.47991667470012, pvalue=2.2620811149354e-05)

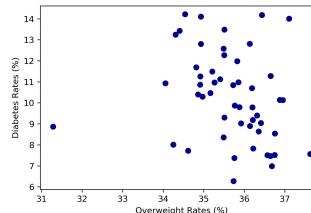
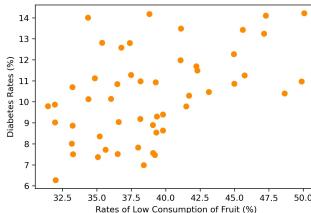
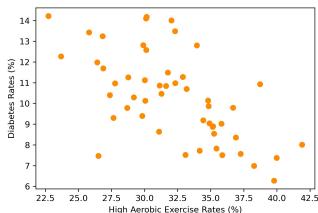


# Predictive Analytics



# Relationship between nutrition, obesity, exercising and diabetes

There were moderate correlations between most of our indicators and diabetes ( $|0.45| < r < |0.71|$ )



Because there were strong correlations among all indicators, we selected one indicator from each group as a predictor in a regression model.

## Obesity

- 01 | Percent of adults who are overweight
- 02 | Percent of adults who are obese

## Exercise

- 01 | Percent of adults who do not exercise
- 02 | Percent of adults who do muscle strengthening workouts 2+ times/week
- 03 | Percent of adults who engage in aerobic exercise, 150 min/week
- 04 | Percent of adults who engage in aerobic exercise, 300
- 05 | Percent of adults who engage in aerobic exercise, 150 min/week + do muscle training 2+ times/week

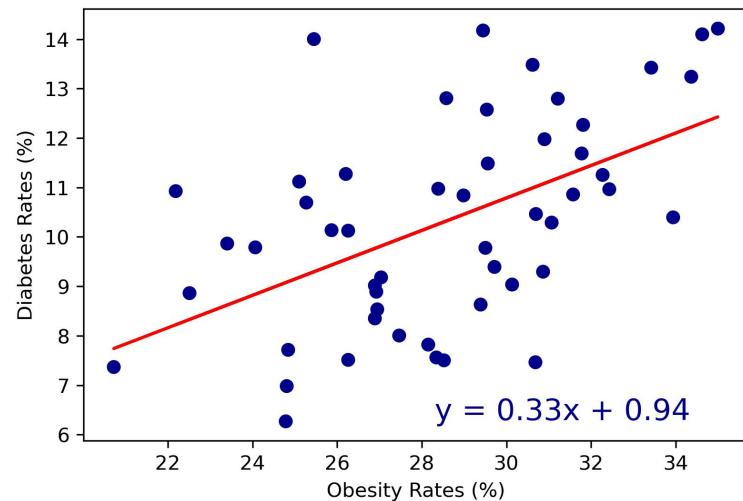
## Nutrition

- 01 | Percent of adults who consume less than one fruit a day
- 02 | Percent of adults who consume less than one vegetable a day

# Predicting Diabetes Based on Obesity

The correlation between diabetes and obesity is moderate ( $r = 0.52$ )

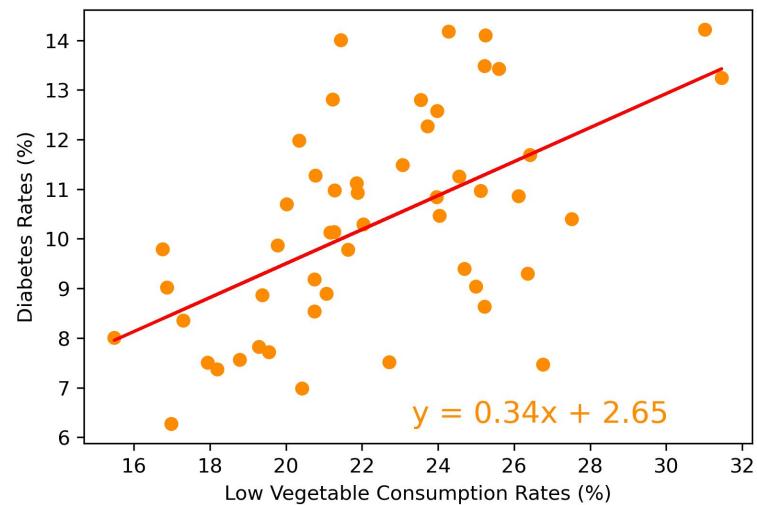
Higher obesity rates in the states in the years 2011-2015 predicts higher diabetes rates in the years 2016-2017



# Predicting Diabetes Based on Nutrition

The correlation between diabetes and low vegetable consumption is moderate ( $r = 0.56$ )

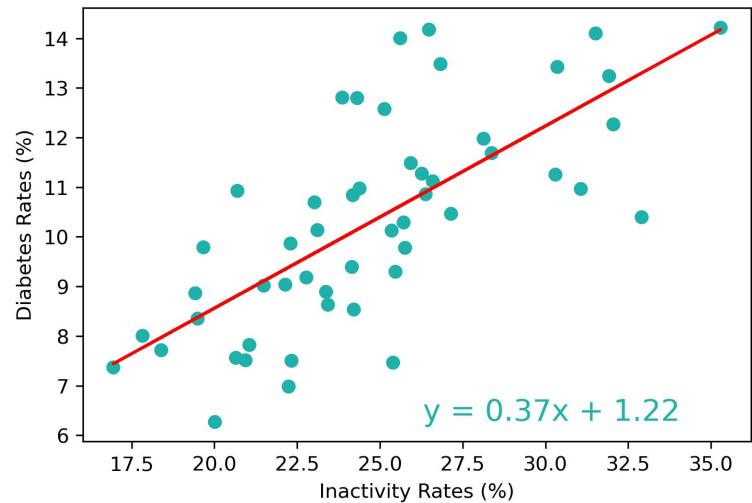
Higher rates of low vegetable consumption in the years 2011-2015 predicts higher diabetes rates in the years 2016-2017



# Predicting Diabetes Based on Exercise

The correlation between diabetes and inactivity is strong ( $r = 0.71$ )

Higher rates of adults who are inactive in the states in the years 2011-2015 predicts higher diabetes rates in the years 2016-2017



# Summary

1

- Nutrition, physical activity and obesity are highly correlated.

2

- Obesity rates in the states explain 27% of the variability in diabetes rates.

3

- Inactivity rates in the states explain 51% of the variability in diabetes rates.

4

- Low vegetable consumption rates in the states shows 31% variability in diabetes rates.



---

Thank you.

