DIABETIC ANALYSIS ON USA COUNTIES

PRESENTED BY

MADHULIKA GUPTA

QUESTIONS ASKED

- Major factors that effect occurrence of diabetes in US population
- Name of factors
  - Obesity
  - Availability of food stores and % of stores that accept SNAP and WIC
  - Income factors(median household income)
  - Availability restaurants(fast food and full service) and % of restaurants that accept SNAP
  - Availability of farmer markets and % that accept the SNAP and WIC

TECHNOLOGIES USED

Back end / cleanup

- Python
- Pandas
- Sklearn

Front End / Data Visualisation

- matplotlib
- Tableau

DATA SOURCE

https://www.ers.usda.gov/data-products/food-environment-atlas/data-access-and-documentation-downloads/#Current%20Version

- Credibility: data published by usda.gov

SEQUENCE OF METHOD ADOPTED FOR THE PROJECT

- Extracting the data from source
- Creating multiple csv's from workbook
- Creating Data frame of each of the csv's
- Selecting and renaming columns
- Merging the worksheets
- Using sklearn to make a linear regression model
  - Use train, test split
  - Use standard scaler
  - Fit and score the model

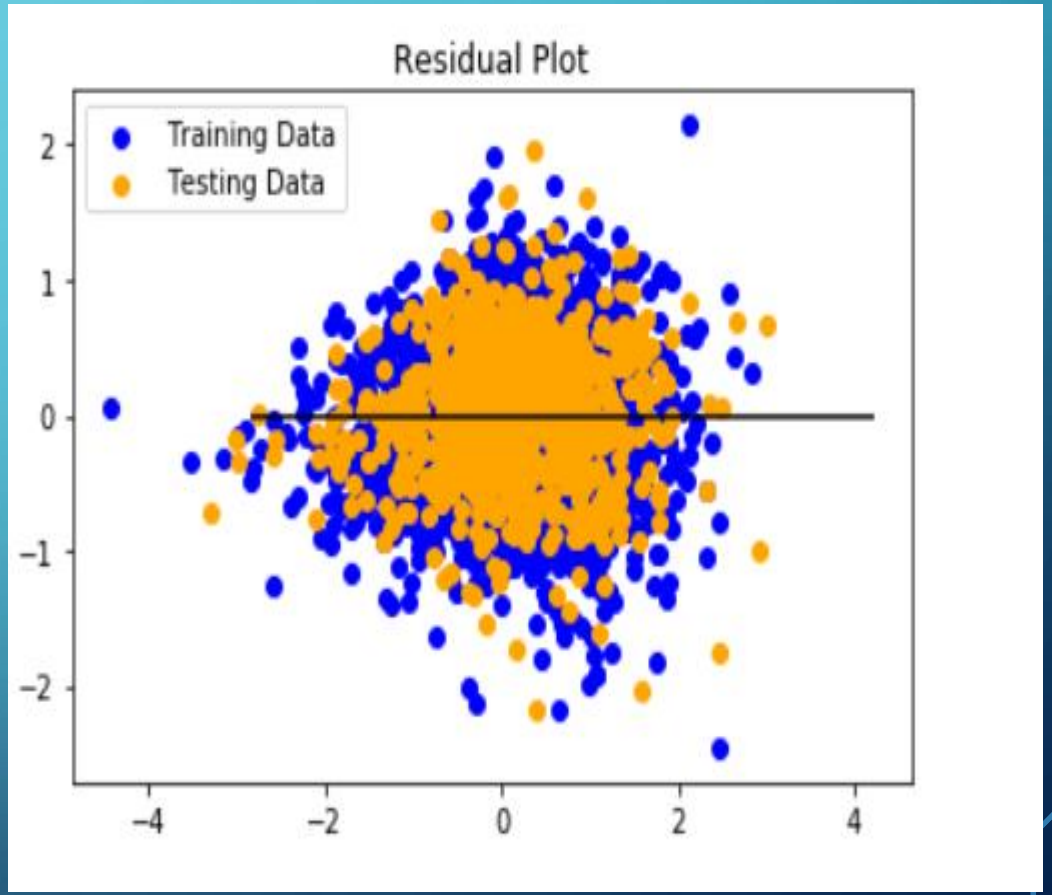
SK-LEARN VALUES

- **Training Score: 0.7023462457014373**
- **Testing Score: 0.7131523555249148**
- **Mean Squared Error (MSE): 0.28**

(mean squared error tells you how close a regression line is to a set of points). A small MSE is better since it implies agreement between the prediction and the reality. MSE is the mean squared difference between your estimate and the data. Smaller MSE generally indicates a better estimate, at the data points in question.

- **R-squared (R2): 0.71**

(R-Squared is a statistical measure of how close are the data fitted to the regression line.



Mean Squared Error (MSE): 0.28343240317564106 R-squared (R2): 0.7131523555249148

PULLING WEIGHTS OF X-FACTORS IN TABLEAU

https://public.tableau.com/profile/madhulika.gupta#!/vizhome/diabetesPredictionAnalysis/ObesityVersusDiabetesperCounty?publish=yes

TAKE AWAYS FROM THE ANALYSIS

- Percentage of population change from 2012 – 2016 has been negligible in most counties , even being negative in some of the.
- Rate of diabetes has risen multiple fold as compared to the population change
- Obesity is a major factor in increase in diabetic rates
- Counts of Stores(including those that accept SNAP) has remained stable all through the years
- The % increase or decrease of fast food restaurants has no direct effect on the diabetes rates.

TAKE AWAYS FROM THE ANALYSIS

- No of markets has remained almost stable(including supercenters, grocery stores , specialized food stores and SNAP authorized markets) and minimal effect on the rate of diabetes.
- No of fast food restaurants have increased in most counties(and could be a contributing factor for increase.
- Diabetic Rate is low where ever median income is high

LESSONS LEARNT

- Obesity and factors leading to Obesity need to be controlled in order to change the rate of diabetes.
- More markets and restaurants authorized to accept SNAP and WIC are required to give access to better quality food.

Thank you