



I. SELECTED TOPIC



SOCIAL MEDIA IMPACT – SUICIDE RATES

UT AUSTIN DATA ANALYTICS & VISUALIZATION
BOOTCAMP – FINAL PROJECT BY :

- QUANG NGUYEN
- TINA BELLON
- IMA RICHBURG
- NENSI DOBARIA
- NAMRATA SHAH



2. REASON FOR TOPIC SELECTION

REASON FOR TOPIC SELECTION

Suicide is a serious societal problem and the result of complex socioeconomic and cultural factors. Even if our analysis should show that social media use only has a minimal impact, we believe any measure to decrease risk and increase suicide prevention is worth pursuing.

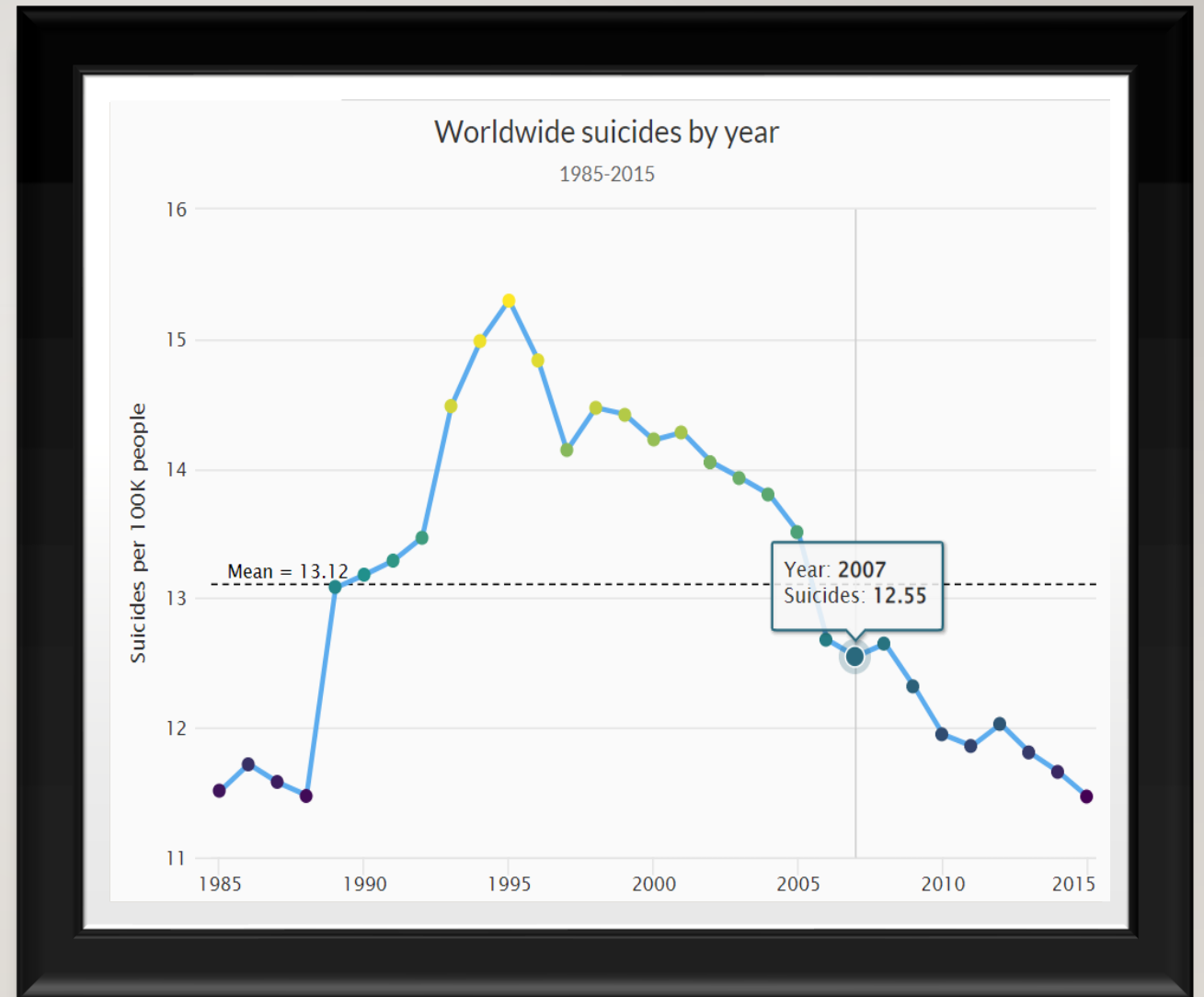


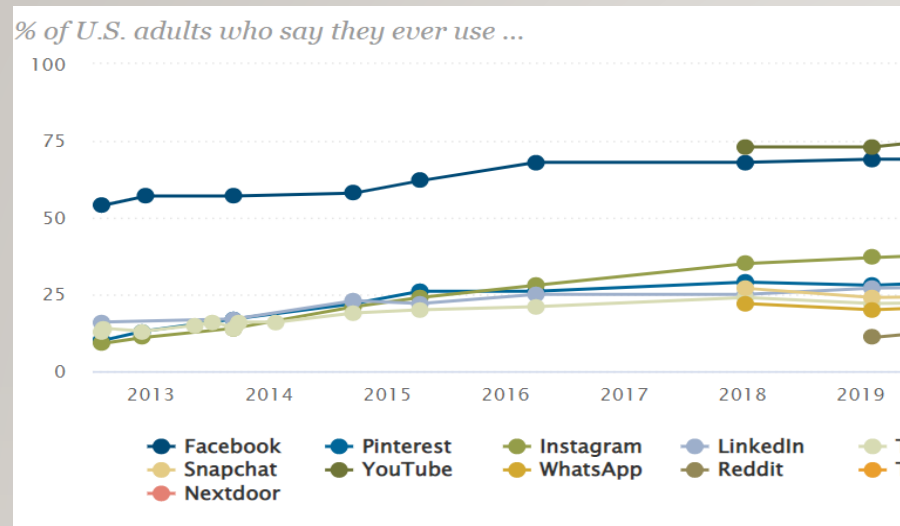


3. DATA SOURCE DESCRIPTION

SUICIDE RATES OVERVIEW 1985 TO 2015 – FROM KAGGLE

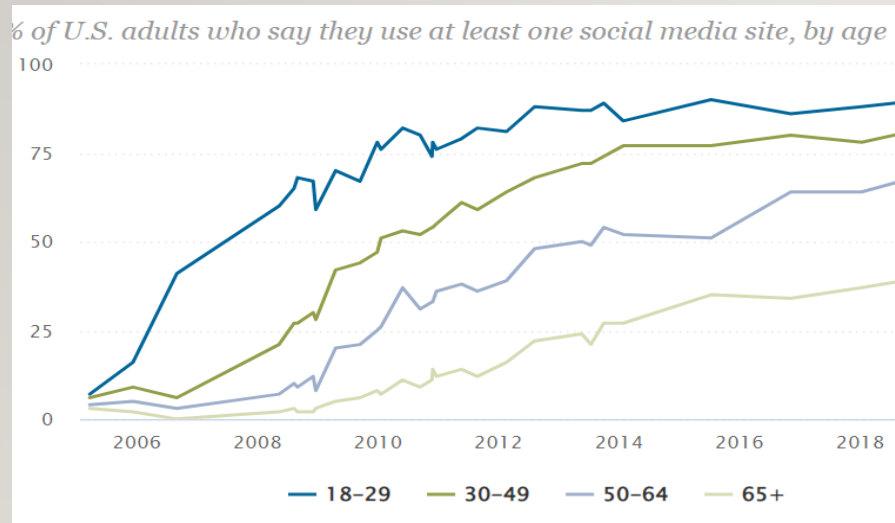
We are using a suicide data set from Kaggle, which lists suicides from around the world by country from 1985 until 2015. The data is split by gender, age and generation, and also lists the respective country's Human Development Index and GDP in a particular year.





SOCIAL MEDIA FACT SHEET – FROM PEW RESEARCH CENTER

For social media usage, we rely on a Pew Research dataset that collected the share of U.S. adults using different forms of social media from 2005 through 2021. The data is broken down by age groups.





4. APPLIED TECHNOLOGIES

TECHNOLOGIES, LANGUAGES, TOOLS, AND ALGORITHMS USED THROUGHOUT THE PROJECT

Languages/toolkits:



Database:



Machine Learning:



Visualization:

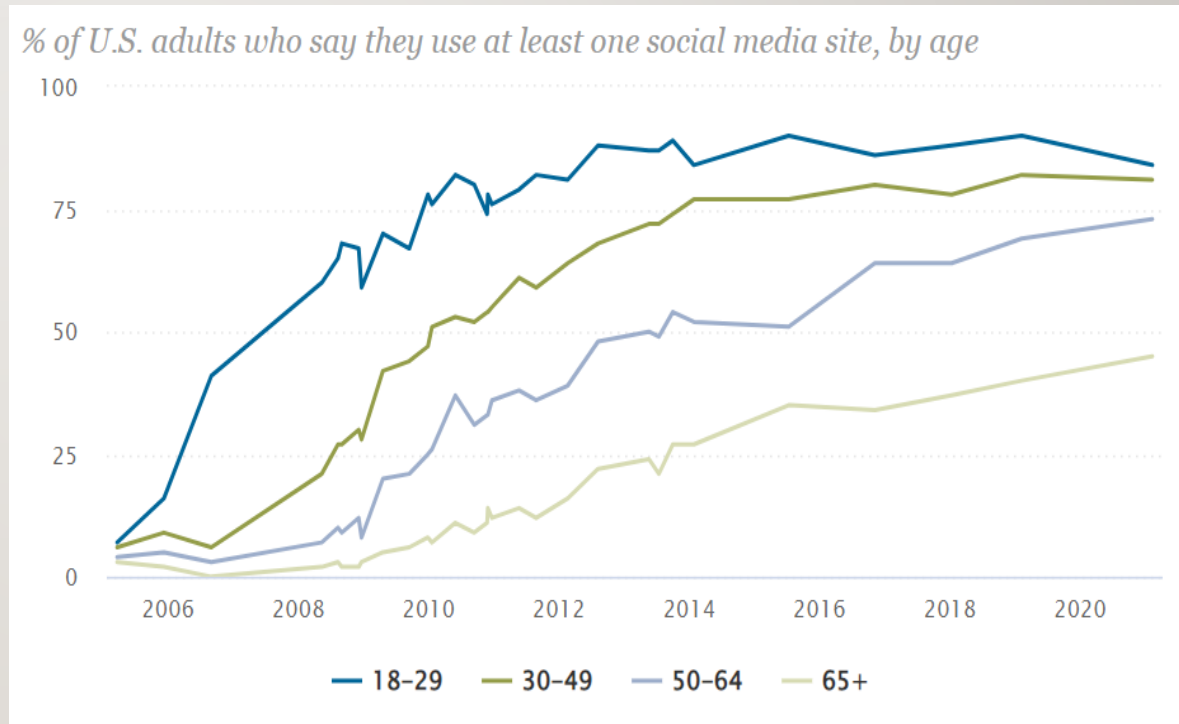




5. ANALYTICAL OBJECTIVES

QUESTIONS WE HOPE TO ANSWER WITH DATA

We wanted to determine whether an uptick in social media use in recent years overlaps with an increase in suicide rates, particularly among younger cohorts who use social media in the greatest numbers. To determine correlation, we plan to visualize U.S. suicide rates and social media use over time.





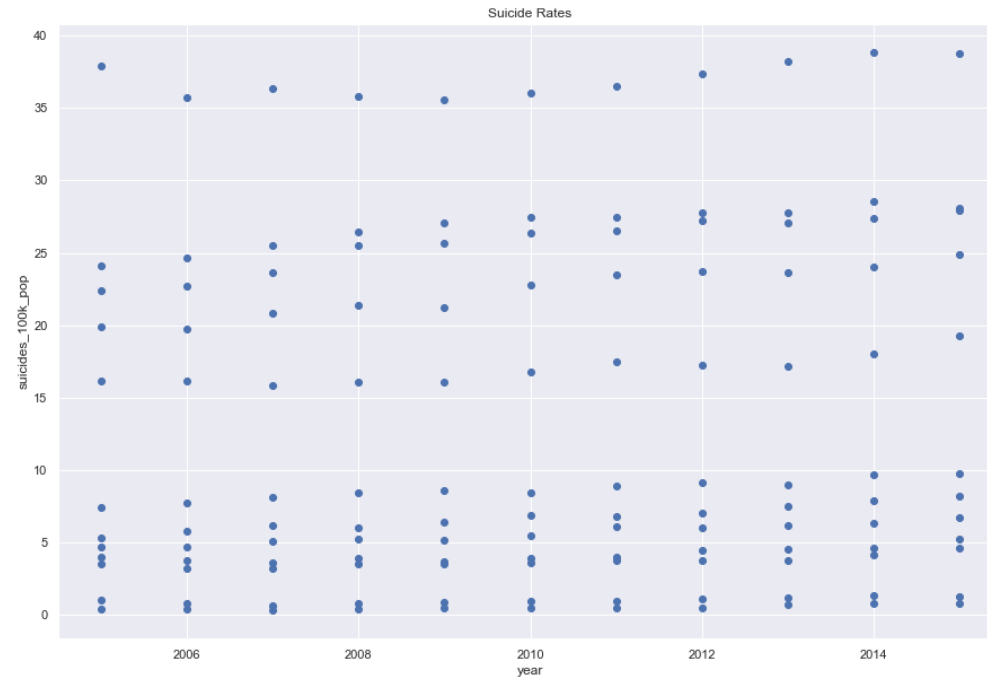
6. DATA EXPLORATION

PANDAS INFO AND DESCRIBE FUNCTIONS

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 27660 entries, 0 to 27659
Data columns (total 12 columns):
 #   Column                Non-Null Count  Dtype
---  -
 0   country                27660 non-null  object
 1   year                   27660 non-null  int64
 2   sex                    27660 non-null  object
 3   age_range              27660 non-null  object
 4   suicides_no            27660 non-null  int64
 5   population              27660 non-null  int64
 6   suicides_100k_pop      27660 non-null  float64
 7   country_year            27660 non-null  object
 8   HDI_for_year            8364 non-null   float64
 9   gdp_for_year            27660 non-null  int64
10   gdp_per_capita          27660 non-null  int64
11   generation              27660 non-null  object
dtypes: float64(2), int64(5), object(5)
memory usage: 2.5+ MB
```

	suicide_year	suicides_no	population	suicides_100k_pop	HDI_for_year	gdp_for_year	gdp_per_capita
count	27820.000000	27820.000000	2.782000e+04	27820.000000	8364.000000	2.782000e+04	27820.000000
mean	2001.258375	242.574407	1.844794e+06	12.816097	0.776601	4.455810e+11	16866.464414
std	8.469055	902.047917	3.911779e+06	18.961511	0.093367	1.453610e+12	18887.576472
min	1985.000000	0.000000	2.780000e+02	0.000000	0.483000	4.691962e+07	251.000000
25%	1995.000000	3.000000	9.749850e+04	0.920000	0.713000	8.985353e+09	3447.000000
50%	2002.000000	25.000000	4.301500e+05	5.990000	0.779000	4.811469e+10	9372.000000
75%	2008.000000	131.000000	1.486143e+06	16.620000	0.855000	2.602024e+11	24874.000000
max	2016.000000	22338.000000	4.380521e+07	224.970000	0.944000	1.812071e+13	126352.000000

VISUAL EXPLORATORY DATA ANALYSIS

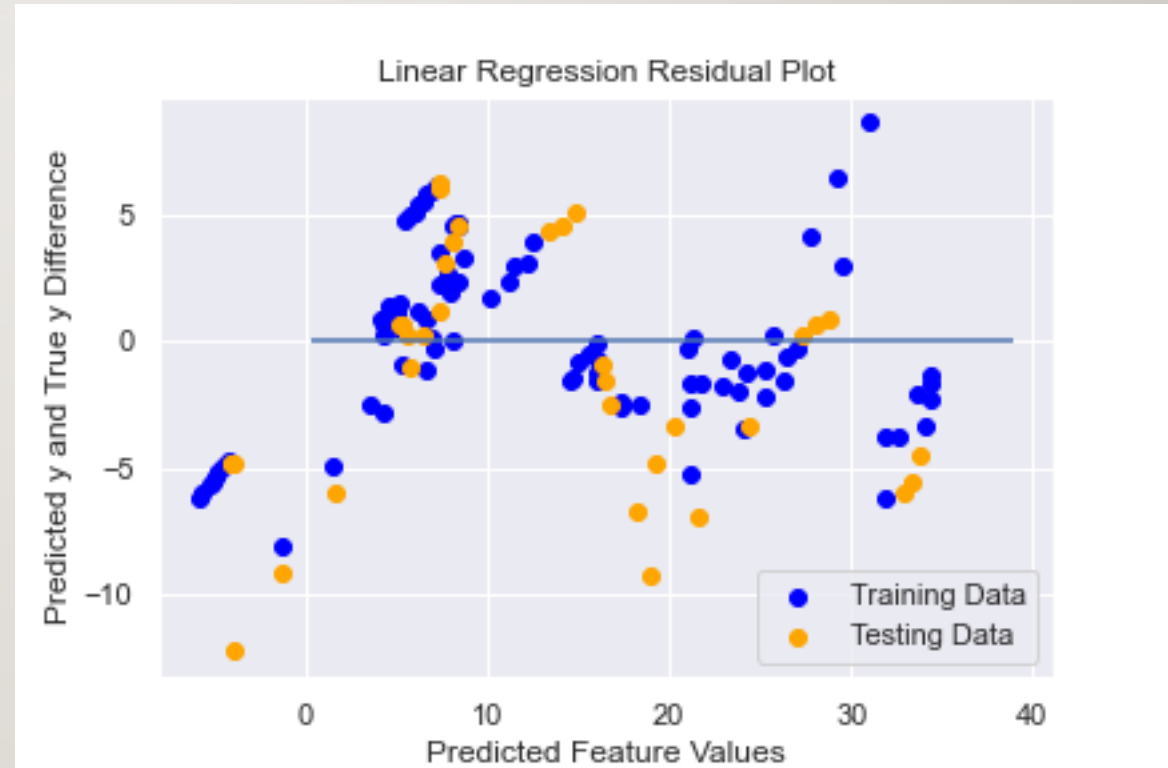




7. DATA ANALYSIS

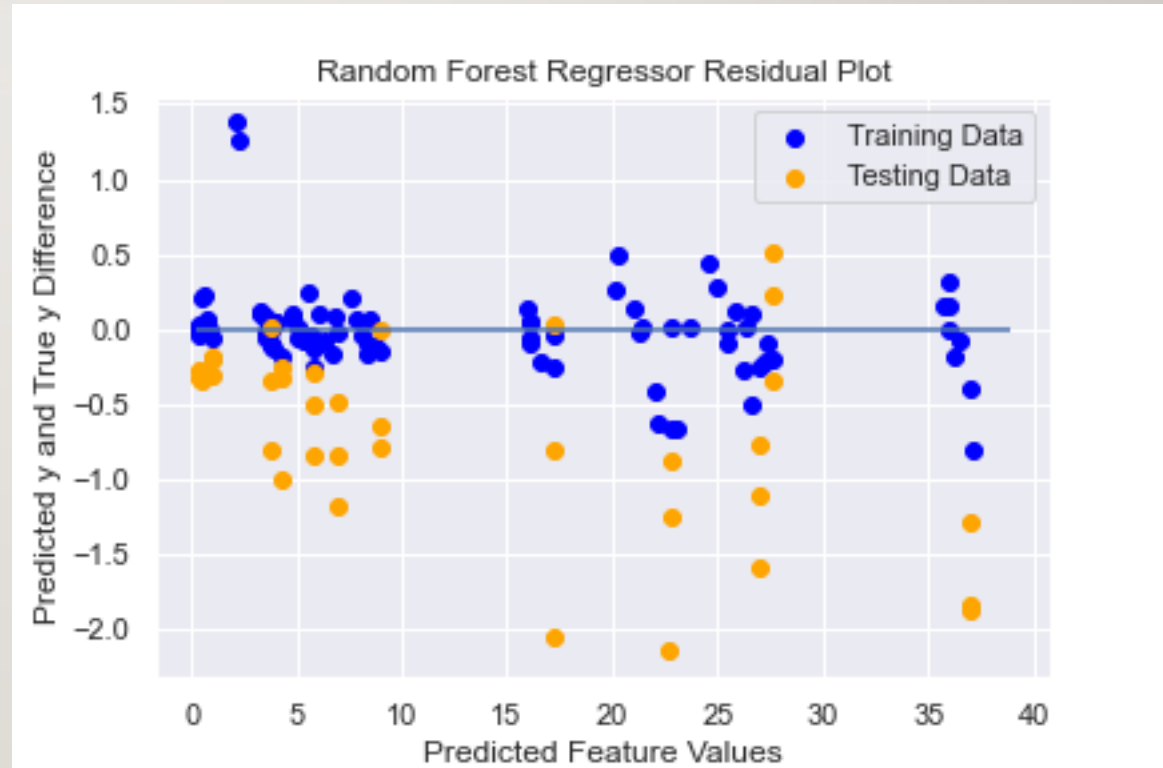
MACHINE LEARNING - LINEAR REGRESSION MODEL

- “Best Fit” line through all data points
- MSE: 24.55
- R-squared: 0.8315
- Training Score: 0.9120
- Testing Score: 0.8315

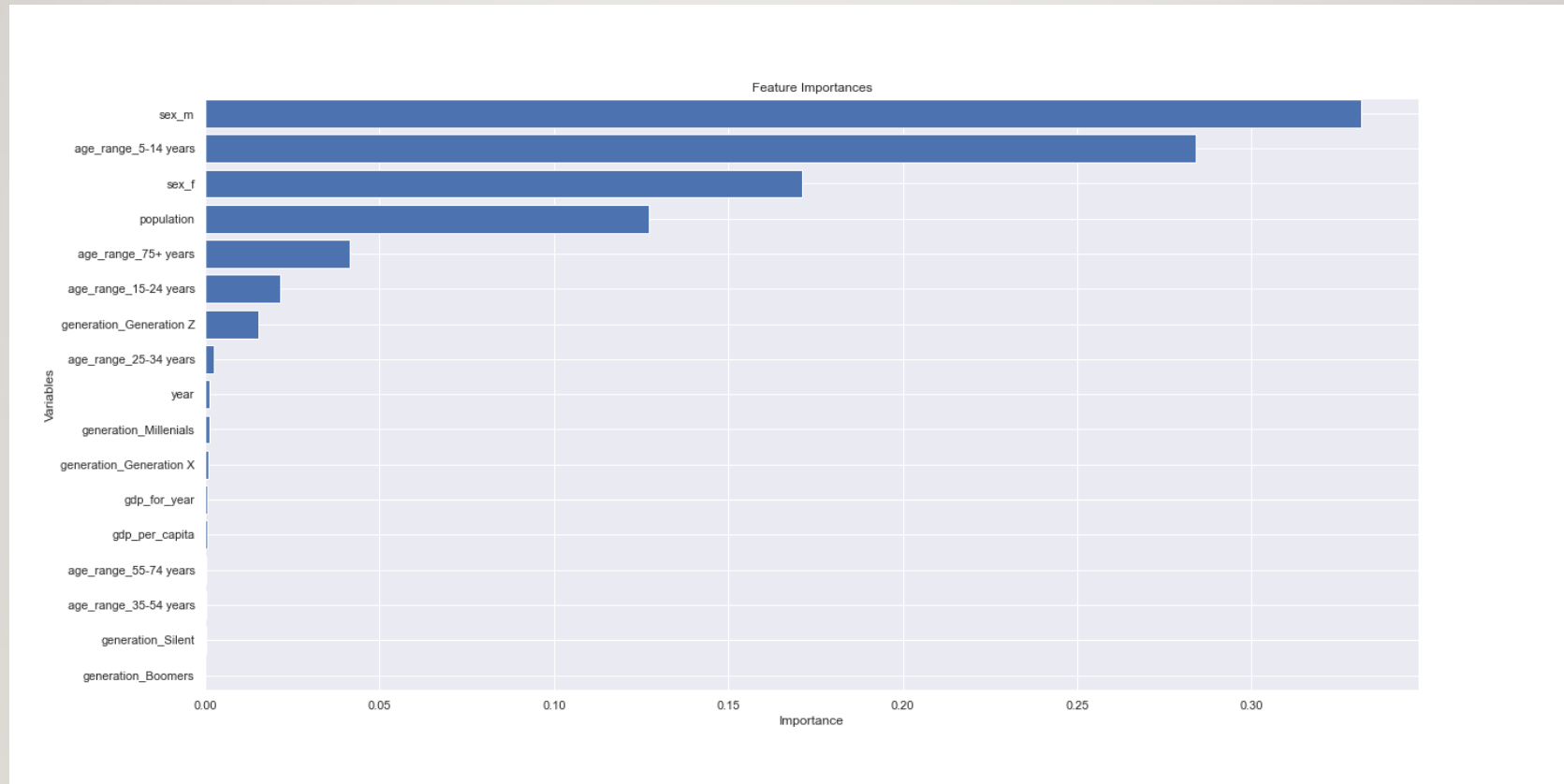


MACHINE LEARNING - RANDOM FOREST REGRESSOR MODEL

- Averages of weaker Decision Trees
- MSE: 0.8826
- R-squared: 0.9939
- Training Score: 0.9994
- Testing Score: 0.9939



RANDOM FOREST REGRESSOR FEATURE IMPORTANCE

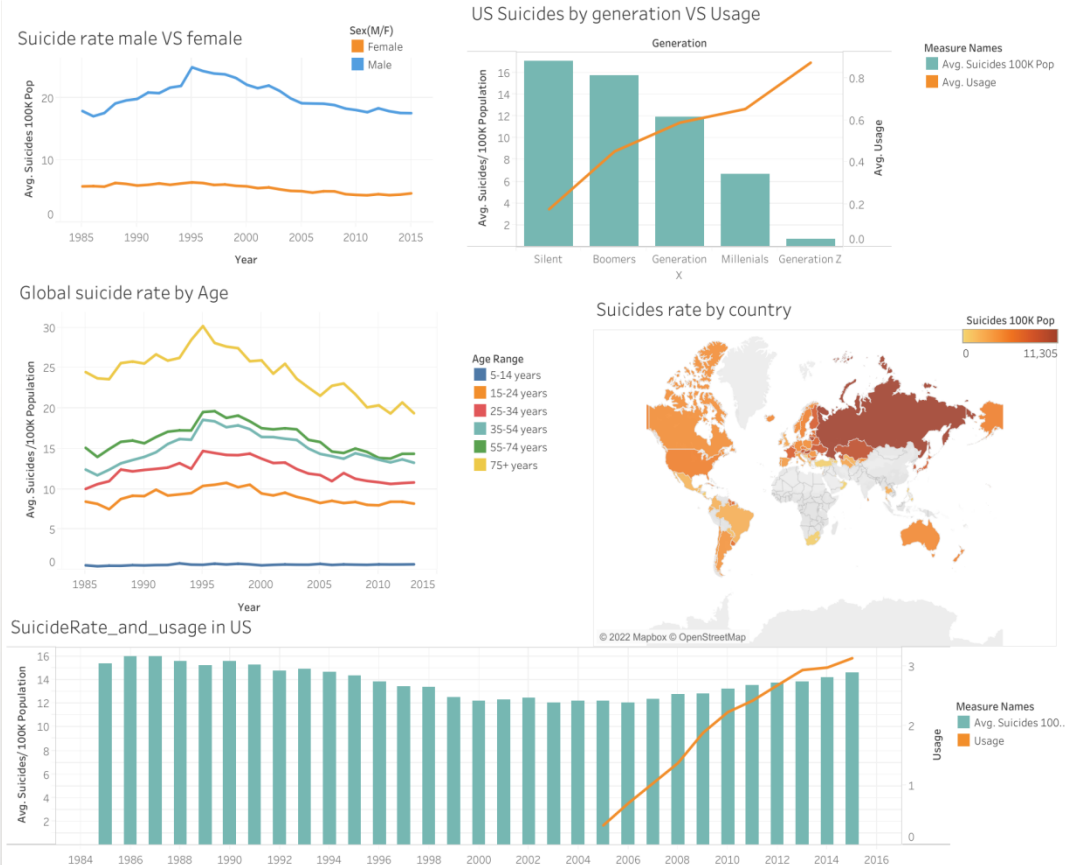





8. RESULTS OF ANALYSIS

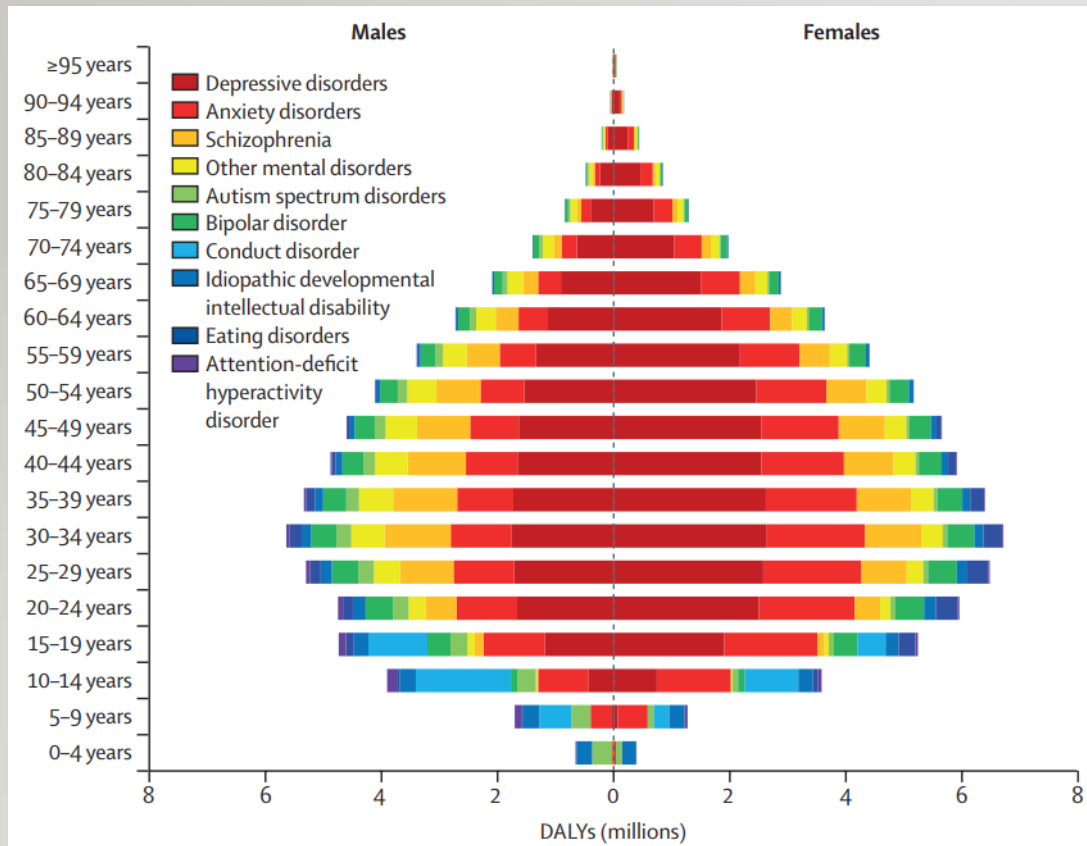
LINK TO OUR VISUALIZATION

- [Tableau Dashboard](#)
- [Final Project Website](#)





9. RECOMMENDATIONS FOR FUTURE ANALYSIS





10. LIMITATIONS



title
Address
city
Washington
Portland
San Francisco
country
USA
phone
415 776 0240
email
emp@emp.com

EmployeeID
EmployeeName
100
101
102

title
Manager
Analyst
Developer

A vibrant, stylized illustration of numerous hands of various colors (white, grey, blue, green, red, yellow) reaching upwards against a solid yellow background. The hands are depicted in a simplified, graphic style with bold outlines. A dark grey horizontal band is positioned across the middle of the image, containing the text.

THANK YOU FOR YOUR
ATTENTION!
