SEARCH ENGINE: https://toolbox.google.com/datasetsearch

1. https://www.kaggle.com/rohankayan/years-of-experience-and-salary-dataset
2. https://www.kaggle.com/eutimiogamboa/years-of-experience-and-salary
3. https://data.world/oecd/gender-wage-gap
4. https://www.kaggle.com/jonavery/incomes-by-career-and-gender
5. https://fraser.stlouisfed.org/title/307/item/6166/toc/335786
6. https://www.ethnicity-facts-figures.service.gov.uk/work-pay-and-benefits/pay-and-income/household-income/latest
7. https://catalog.data.gov/dataset/demographics-ec791

https://catalog.data.gov/dataset?q=salary+gender&sort=views_recent+desc&ext_location=&ext_bbox=&ext_prev_extent=-142.03125%2C8.754794702435618%2C-59.0625%2C61.77312286453146
New goal: predict salary
Features (possible combination of): race, gender, age, company, marital status

!!!: https://www.bls.gov/cps/earnings.htm

- Table 3: by state
- Table 7: marital status
- Table 13: race
    - Not sure how it's different from table 16
- Table 14: education level time series
- Table 15: age time series
- **tables 11-16 are the percentage ones (main interest)**

!!:
https://www.enigma.com/blog/exploring-social-issues-through-public-data-the-gender-wage-gap

NEW DATASETS:

- https://query.data.world/s/upd25opte27gwfxswfr7bfuhyus23d (I think this one would be best to work with)
    - (Comments: An idea I had for this dataset would be to predict annual bonuses based on company, category of work and years of experience etc.)
        - What is the motivation/impact, though?

- https://query.data.world/s/hexs5527qevwl4z4h3ffalxkkcsccj
    - (Comments: An idea for this dataset would be to predict total compensation based on which company/ industry an individual works in etc.)
        - What is the motivation/impact, though?

Link	Motivation	Independent Vars	Dependent Var	Notes
https://www.kaggle.com/ronitf/heart-disease-uci	Predicting heart disease – help patients with chest pain determine if they really have heart disease	All other columns	Target	only concern is that this is p much the project on kaggle and also only 303 rows
https://www.kaggle.com/karangadiya/fifa19 (scraped from https://sofifa.com/players)	Predict salary of FIFA world cup soccer players (are they overpaid/underpaid? – compare to their actual salary Could that money be used for other things instead?)	Age, Nationality, Overall rating, potential rating, value? (this might have a correlation with wage), international reputation, skill moves, body type	Wage	Would be nice to have how many goals/assists each player has had both in world cup and in their clubs
https://www.kaggle.com/datasf/san-francisco bikeshare_trips	Predict number of trips (?) based on time of day and location? Goal: set up safety things for bikers to prevent accidents when more bikers are out	Start station name, end station name (or zip code instead of previous two), start date time, end date time, (duration?),	An aggregated stat for number of trips per day?	Too similar to data 100's bikesharing project? Also hesitant about this one
https://www.kaggle.com/russellyates88/suicide-rates-overview-1985-to-2016	Suicide prevention	Year, country, age/generation, sex, population, gdp per capita, HDI (human development index) if available	suicides_no OR suicides/100k pop	HDI is missing from some rows
https://www.kaggle.com/lislejoem/us-minimum-wage-by-state-fr	Predict minimum wage – can use trajectory to compare to cost	State, year, CPI.average, High.2018, Low.2018	High.value OR low.value	Not sure if enough data if we split by state (unless we don't

om-1968-to-2017	of living and impacts things like social security etc in future			need to?) also maybe not enough columns to use to predict?
https://www.kaggle.com/mohansacharya/graduate-admissions	Predict the chance of students' admission to graduate universities	All other columns	Chance of Admission	SerialNo is just the IDs
https://healthdata.gov/dataset/500-cities-census-tract-level-data-gis-friendly-format-2019-release	Predict colorectal cancer screening outcome	all teeth lost, dental visits, mammograms, Pap tests, core preventive services among older adults, and sleep less than 7 hours and location	Screening outcome	N/A
https://www.kaggle.com/worldbank/world-development-indicators				

http://apps.who.int/healthinfo/statistics/mortality/whodpms/

- Instructions: https://www.who.int/mental_health/suicide-prevention/extraction_suicide_statistics.pdf?ua=1

HDI data from: http://hdr.undp.org/en/indicators/137506

https://ourworldindata.org/suicide

New Data

- CDC List of Trends Table: https://www.cdc.gov/nchs/data/hus/hus18.pdf
- Percentage of U.S. population without health care visits in the past 12 months from 1997 to 2017, by gender: https://www.statista.com/statistics/189319/us-population-without-health-care-visits-by-gender-since-1997/
- Respondent-assessed fair-poor health status, by selected characteristics: United States, selected years 1991–2017: https://www.cdc.gov/nchs/hus/contents2018.htm#Table_016
    - Backup maybe?
- ***Drug overdose death rates, by drug type, sex, age, race, and Hispanic origin: United States, selected years 1999–2017: https://www.cdc.gov/nchs/hus/contents2018.htm#Table_008***
- *Selected health conditions and risk factors, by age: United States, selected years 1988-1994 through 2015-2016: https://www.cdc.gov/nchs/hus/contents2018.htm#Table_021*
- Delay or nonreceipt of needed medical care, nonreceipt of needed prescription drugs, or nonreceipt of needed dental care during the past 12 months due to cost, by selected characteristics: United States, selected years 1997–2017: https://www.cdc.gov/nchs/hus/contents2018.htm#Table_029
- Health care visits to doctor offices, emergency departments, and home visits within the past 12 months, by selected characteristics: United States, selected years 1997–2017: https://www.cdc.gov/nchs/hus/contents2018.htm#Table_030
- Emergency department visits within the past 12 months among children under age 18, by selected characteristics: United States, selected years 1997–2017: https://www.cdc.gov/nchs/hus/contents2018.htm#Table_035
- Emergency department visits within the past 12 months among adults aged 18 and over, by selected characteristics: United States, selected years 1997–2017: https://www.cdc.gov/nchs/hus/contents2018.htm#Table_036
- Gross domestic product, national health expenditures, per capita amounts, percent distribution, and average annual percent change: United States, selected years 1960–2017: https://www.cdc.gov/nchs/hus/contents2018.htm#Table_042
- National health expenditures, average annual percent change, and percent distribution, by type of expenditure: United States, selected years 1960–2017: https://www.cdc.gov/nchs/hus/contents2018.htm#Table_043
    - 10 year interval instead of yearly

- Personal health care expenditures, by source of funds and type of expenditure: United States, selected years 1960–2017: https://www.cdc.gov/nchs/hus/contents2018.htm#Table_044 (All personal health care expenditures)
    - 10 year interval instead of yearly
- 2018-2019 food security index: https://foodsecurityindex.eiu.com/Index
- Food sustainability index (not sure what year?): http://foodsustainability.eiu.com/heat-map/ by country
- Food supply - Crops Primary Equivalent: http://www.fao.org/faostat/en/#data/CC
    - Main page: http://www.fao.org/faostat/en/#data
    - Led to here from: https://ourworldindata.org/food-per-person#data-sources
- Inequality of Food Consumption – Coefficient of variation (and skewness) of habitual caloric consumption distribution since 1990: http://web.archive.org/web/20141108183923/http://www.fao.org:80/economic/ess/ess-fs/fs-data/en/
    - Pretty solid except they have spotty nan values
- Percentage of adults in the U.S. with depression from 2013 to 2016, by age and gender: https://cdn.statista.com/statistics/815321/depression-among-adults-us-by-age-and-gender/
    - Only issue is that it's only 2013-2016 and our data goes back to like 1985
- Number of adults with mood disorders in the U.S. in 2008-2012, by age group: https://www.statista.com/statistics/451768/mood-disorder-number-among-adults-in-the-us-by-age-group/
    - Same comment as above (limited years)
- **Percentage of people in the U.S. who suffered from depression from 1990 to 2017, by gender: https://www.statista.com/statistics/979898/percentage-of-people-with-depression-us-by-gender/**
    - Solid (we'd have to find an equivalent for all the other countries we do too though)
- Percentage of adults in the U.S. aged 65 years and older with clinically relevant depressive symptoms from 1998 to 2014, by gender: https://www.statista.com/statistics/726506/depressive-symptoms-among-seniors-united-states-by-gender/
    - Above one is prob better (bc of greater year span coverage?)
- Quality of life index VS level of happiness in 2017 across different countries: https://zenodo.org/record/1470818#.XfCDzi_Myu4
    - Seems p cool but only data for 2017 (our data only goes to 2015)
- Income level at which money won't make you happier in the United States in 2010, state-by-state comparison: https://static1.statista.com/statistics/319651/happiness-benchmark-in-the-us/
    - Also cool but our data isn't granular at state level :(and only data for 2010
- Different mental health and disorder types https://ourworldindata.org/mental-health

    - Cool but we’d have to web scrape and only data for 2017
- Percentage of worldwide population that had depression from 1990 to 2017
- https://www.statista.com/study/65989/mental-health-worldwide/