**Pyramid Group**

|  |  |
| --- | --- |
| **FINAL IDEA** | **Give your information in this column** |
| TITLE | The Effect of Youth Tobacco Consumption on Mortality Rate |
| Research question | If there exists a correlation between youth tobacco consumption and some types of mortality such as cancer mortality, heart disease mortality, drug overdose mortality rate etc. |
| Data source: exact full web link | Primary:  [Youth Tobacco Survey (YTS) Data](https://chronicdata.cdc.gov/Survey-Data/Youth-Tobacco-Survey-YTS-Data/4juz-x2tp)  Secondary:  [Cancer Mortality by State](https://www.cdc.gov/nchs/pressroom/sosmap/cancer_mortality/cancer.htm)  [Heart Disease Mortality by State](https://www.cdc.gov/nchs/pressroom/sosmap/heart_disease_mortality/heart_disease.htm)  [Drug Overdose Mortality by State](https://www.cdc.gov/nchs/pressroom/sosmap/drug_poisoning_mortality/drug_poisoning.htm)  [Stroke Mortality by State](https://www.cdc.gov/nchs/pressroom/sosmap/stroke_mortality/stroke.htm)  We would merge them on the state feature. |
| Is data downloadable? What format? | Yes. In csv format. |
| List of independent variables (7-10) | Year, Longitude, Latitude, State, Smoke Type, Gender, Education Level |
| List of dependent variables (5-7) | Smoke Rate, Cancer Mortality Rate, Heart Disease Mortality Rate, Drug Overdose Mortality Rate, Stroke Mortality Rate |
| # of years data, period | 5 years of data, 2013 - 2017 |
| Data types (integer, numeric, alphanumeric, etc)  Data scales (ordinal, nominal, ratio, interval) | Mixed (include integer, float, numeric, characters)  Ordinal, eg. mortality rate level - low, middle, high  Ratio, eg. Numeric mortality rate - 0.112 |
| Data size (# of years X # of variables X # of rows) = give formula and total size | 5 years \* 12 variables \* 10601 rows = 636060 |
| Drill down/multidimensional analysis capability – Y/N?  Example | Yes, we can have a country level perspective first, then drill down into state level. We can check, for example, would the correlation coefficient vary among states if there does exist some correlation between Tobacco consumption and mortality rate |
| Certify – original idea – Y/N? | Yes. This is our original idea. |