|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Hypothesis** | **Question** | **Analytic type** | **Variable** | **Chart** |
| Smoking rate is associated with different demographic segmentations such as gender, education and smoking type | Do different gender, education level, and smoking type have the same smoking rate? | Descriptive | Gender, Education, TopicDesc, Smoke\_Rate | Stacked bar chart |
|  | What are the top 5 states with the highest smoke rate? | Descriptive | STATE, Smoke\_Rate | Map with symbol |
| There are outliers in mortality rates (cancer, heart disease, drug overdose, stroke) | Is there any mortality rate record unusually high/low? | Descriptive | CANCER\_RATE, HEART\_DISEASE\_RATE, OVERDOSE\_RATE, SREOKE\_RATE | Box and whisker plots |
| The standard deviation of smoking rates follows a normal distribution. | Is the standard deviation of the smoking rate Gaussian distributed? | Descriptive | Smoke\_Rate | Histogram |
|  | Is there any difference between the middle school’s smoking rate distribution and the high school’s smoking rate distribution? | Descriptive | Education, Smoke\_Rate | Faceted Histogram |
| The total mortality rate varies in different smoke types (cessation, cigarette, smokeless tobacco) | Which smoke type has the highest mortality rate in 2014? | Descriptive | TopicDesc,CANCER\_RATE, HEART\_DISEASE\_RATE, OVERDOSE\_RATE, SREOKE\_RATE | Horizontal Bar Chart |
|  | Which smoke type has the highest mortality rate in 2015? | Descriptive | TopicDesc,CANCER\_RATE, HEART\_DISEASE\_RATE, OVERDOSE\_RATE, SREOKE\_RATE | Horizontal Bar Chart |
|  | Which smoke type has the highest mortality rate in 2016? | Descriptive | TopicDesc,CANCER\_RATE, HEART\_DISEASE\_RATE, OVERDOSE\_RATE, SREOKE\_RATE | Horizontal Bar Chart |
|  | Which smoke type has the highest mortality rate in 2017? | Descriptive | TopicDesc,CANCER\_RATE, HEART\_DISEASE\_RATE, OVERDOSE\_RATE, SREOKE\_RATE | Horizontal Bar Chart |
|  | What will happen to the heart disease rate of smoke types of cessation in NJ after 2017? | Predictive | YEAR, STATE,  TopicDesc,  HEART\_DISEASE\_RATE | Continuous line chart and forecast tool from Tableau |
|  | What will happen to the overdose rate of smoke types of cessation in NJ after 2017? | Predictive | YEAR, STATE,  TopicDesc,  OVERDOSE\_RATE | Continuous line chart and forecast tool from Tableau |
|  | What will happen to the stroke rate of smoke types of cessation in NJ after 2017? | Predictive | YEAR, STATE,  TopicDesc,  STROKE\_DEATHS | Continuous line chart and forecast tool from Tableau |
|  | What will happen to the cancer rate of smoke types of cessation in NJ after 2017? | Predictive | YEAR, STATE,  TopicDesc,  CANCER\_RATE | Continuous line chart and forecast tool from Tableau |
|  | What will happen to the cancer rate of smoke types of cigarette use in NJ after 2017? | Predictive | YEAR, STATE,  TopicDesc,  CANCER\_RATE | Continuous line chart and forecast tool from Tableau |
| In detail, cessation, smoked cigarettes or smokeless tobacco correlate with mortality rate of stroke, heart attack, overdose, and cancer. | What is the distribution of each moarlity rate based on three smoke types? | Descriptive | TopicDesc,  CANCER\_RATE, HEART\_DISEASE\_RATE, OVERDOSE\_RATE, SREOKE\_RATE | 9 Scatter plots in a dashboard |
| The smoking rates do not vary in different smoking types (cessation, cigarette, smokeless tobacco). | Is there any significant difference among the smoking rates for different smoking types? | Predictive | TopicDesc,CANCER\_RATE, HEART\_DISEASE\_RATE, OVERDOSE\_RATE, SREOKE\_RATE | NA - it’s a statistical model.  Model: ANOVA test |
| Smoke rate is positively correlated with mortality rate | What is the trend of smoke rate and different mortality rates from 2014 to 2017? | Descriptive | YEAR, Smoke\_Rate, CANCER\_RATE, HEART\_DISEASE\_RATE, OVERDOSE\_RATE, SREOKE\_RATE | Continuous Line Chart |
|  | For each state student, what is the trend of smoke rate and total mortality rates after 2017 for different genders? | Predictive | YEAR, Smoke\_Rate,Gender, CANCER\_RATE, HEART\_DISEASE\_RATE, OVERDOSE\_RATE, SREOKE\_RATE | Continuous Line Chart and forecast tool from Tableau |
| The death rate is not associated with sample size | Is there an association between death counts and sample sizes? | Descriptive | Sample\_Size, HEART\_DISEASE\_DEATHS, OVERDOSE\_DEATH, STROKE\_DEATHS, CANCER\_DEATHS | Area chart |
| Gender may be an important feature and associated with mortality rate of stroke, heart attack, overdose, and cancer. | What is the smoking rate and mortality rate difference just based on the gender difference? | Descriptive | Gender, Smoke\_Rate, CANCER\_RATE, HEART\_DISEASE\_RATE, OVERDOSE\_RATE, SREOKE\_RATE | Highlight tables |
| There exists a random forest model that can accurately predict high or low cancer mortality rate. | Given all independent variables, can we predict whether a states’ smokers have a high or low cancer mortality rate. | Predictive | YEAR，STATE，TopicDes， MeasureDesc， Smoke\_Rate， Data\_Value\_Std\_Err, Sample\_Size, Gender, Education, HEART\_DISEASE\_RATE, HEART\_DISEASE\_DEATHS, OVERDOSE\_RATE, OVERDOSE\_DEATH, SREOKE\_RATE, STROKE\_DEATHS, CANCER\_RATE, CANCER\_DEATHS | No charts - it’s a machine learning model  Model: Random forest  (We may also try SVM, XGBoost, logistic regressions etc.) |