RITDataBI

Group Regression Project

Due 3-3-2020

You are hired by American Cancer Society (<https://www.cancer.org/>) to write a white paper exploring the factors related to cancer deaths in the US. They will use your analysis to identify regions (and associated partners) for cancer interventions across the US. The dataset provided includes information about all zipcodes in the United States.

Your task: Create a model of cancer incidence using the factors in this dataset. Write an executive summary of cancer factors in America, which can be read by non-statistical business decision makers.

Specifically answer following questions:

1. Which regions of the country are most prone to cancer?
2. Conduct hypothesis tests to determine if the following lead to **more** cancer incidence Use alpha of 5%.
   1. Recent trend stable vs falling (Column M)
   2. Study Count (Column C). Create a new variable to recode this As High vs Low and do the hypothesis test. Break the two groups at the mean study count.
3. Create a 4 level indicator variable for Median Income (column G) and conduct an ANOVA comparing the four groups. Four levels might be Very low, low, high, very high or similar
4. Do a correlation analysis to see which factors are most highly correlated with cancer incidence rate (and which are correlated with each other!)
5. Create an actionable regression model for incidence rate. It should explain as much of the variability as possible but only identify the most important factors.
6. EXPLAIN your findings so that a non-technical person can understand!

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **GROUP 1**   |  |  |  |  | | --- | --- | --- | --- | | |  | | --- | | Marshall Hurst | | Condy Kan | | Noah Blankenship | | | **GROUP 2**   |  |  |  |  |  | | --- | --- | --- | --- | --- | | |  | | --- | | Jeet Patel | | Derek Piotrowski | | Krista Stanislow | |  | |  |  | |  |  | |
| **GROUP 3**   |  | | --- | | Joshua Bhatti | | Ruidong Li | | Nan Meng | | **GROUP 4**   |  |  | | --- | --- | | Amanda Nesheiwat | | | Christina Maher | | | Braedon Williams | | | |  | | --- | |  | |  | | |  |  | | | **GROUP 6** |  | | |
| **GROUP 5**   |  | | --- | | Mitej Patel | | Emilee McHugh | | Mike Sandway  Zach Wiser | | |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | | |  | | --- | | Katey Sackett | | Ninad Sunil Godambe | | Amelia Lombardo | | Austin Vyskocil | |  | |

**GROUP 7**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| |  | | --- | | Mark Saile | | Victor Madruga Cavalcanti da Silva | | Kevin Yirinec | | Mohammad Akbar Nory | |  |