

DATA 622: Machine Learning and Big Data: HW4 (Final Project)

Gabriel Campos

Last edited May 17, 2024

Assignment Description

Exploratory analysis and essay

Assignment

1. Choose a dataset You get to decide which dataset you want to work on. The data set must be different from the ones used in previous homeworks You can work on a problem from your job, or something you are interested in. You may also obtain a dataset from sites such as Kaggle, Data.Gov, Census Bureau, USGS or other open data portals.
2. Select one of the methodologies studied in weeks 1-10, and another methodology from weeks 11-15 to apply in the new dataset selected.
3. To complete this task:.
 - a. Describe the problem you are trying to solve.
 - b. Describe your datasets and what you did to prepare the data for analysis.
 - c. Methodologies you used for analyzing the data
 - d. What's the purpose of the analysis performed
 - e. Make your conclusions from your analysis. Please be sure to address the business impact (it could be of any domain) of your solution.

Deliverable

1. Your final presentation (essay or video) should include:
 1. The traditional R file or Python file and essay,
 2. An Essay (minimum 500 word document) or Video (5 to 8 minutes recording) Include the execution and explanation of your code. The video can be recorded on any platform of your choice (Youtube, Free Cam).

Libraries

```
library(Amelia)
library(car)
library(caret)
library(corrplot)
library(Cubist)
library(DataExplorer)
library(dplyr)
library(e1071)
library(earth)
library(forcats)
library(forecast)
library(fpp3)
library(gbm)
library(ggplot2)
library(kableExtra)
library(MASS)
library(mice)
library(mlbench)
library(party)
library(randomForest)
library(RANN)
library(RColorBrewer)
library(readxl)
library(rpart)
library(rpart.plot)
library(summarytools)
library(tidyr)
library(VIM)
library(earth)
library(randomForest)
```

Overview

National Center for Health Statistics
Mortality Data on CDC WONDER

All Ages Deaths by Underlying Cause

Underlying Cause of Death

2018-2022: Underlying Cause of Death by Single-Race Categories

1999-2020: Underlying Cause of Death by Bridged-Race Categories

1968-2016: Compressed Mortality

The mortality data available on CDC WONDER are national mortality and population data produced by National Center for Health Statistics (NCHS) at the Centers for Disease Control and Prevention (CDC). Mortality information is collected by state registries and provided to the National Vital Statistics System. Data are based on death certificates for U.S. residents. Each death certificate contains a single underlying cause of death, and demographic data. The number of deaths and death rates can be obtained by place of residence (United States national, state, and county when available), age group, race, Hispanic ethnicity, gender, and cause of death (4-digit ICD-10 codes, 113 selected causes of death, 130 selected causes of death for infants, and categories for injury intent and mechanism, or drug / alcohol induced causes of death, when available). For more information, refer to [National Vital Statistics System - Mortality Data](#).

Page last reviewed: April 26, 2024
Content source: CDC WONDER

Underlying Cause of Death, 1999-2020 Request
Deaths occurring through 2020

[Home](#)
[Tutorial](#)
[Help](#)
[About](#)

Processing Request, please wait for results...

[Indicator/View Details of Death Rate](#)
[Default Measurement](#)
[Other Data Access](#)
[Data User Registration](#)
[Help Us Use JHUP](#)

[Save](#)
[Reset](#)

Make all desired selections and then click any **Send** button one time to send your request.

1. Organize table layout

[Send](#)
[Help](#)

Group Results By:
 And By:
 And By:
 And By:
 And By:

Measures:
☐ Deaths ☐ Population ☐ Crude Rate
 For crude rate: ☐ 95% Confidence Interval ☐ Standard Error
☐ Age Adjusted Rate ☐ 95% Confidence Interval ☐ Standard Error
☐ Percent of Total Deaths
 Title:

☐ **Additional Data Options** Click "+" for non-standard age adjusted rates and other options.

2. Select location(s)

[Send](#)
[Help](#)

Click a button to select locations by State, Census Region or HHS Region.

Status:

Browse or search to find items in the States/Region/then highlight the items to use for this request.
 (The Currently selected box displays all current request items.)

[Finder Tool Help](#) [Advanced Finder Options](#)

[Browse](#)
[Search](#)
[Delete](#)

States

☐ 01 (Alabama)
☐ 02 (Alaska)
☐ 04 (Arizona)
☐ 05 (Arkansas)
☐ 06 (California)
☐ 08 (Colorado)
☐ 09 (Connecticut)
☐ 10 (Delaware)

Currently selected:
☐ All
☐ The United States

Browse the list by opening and closing items.
 Use Ctrl+Click to multiply select, Shift+Click for a range.

Click between:
 2008 Urbanization ☐
 2008 Urbanization ☐

2013 Urbanization
☐ All Categories
☐ Large Central Metro
☐ Large Fringe Metro
☐ Medium Metro
☐ Small Metro
☐ Microcosplan (Nonmetro)
☐ NonCore (Nonmetro)

3. Select demographics

[Send](#)
[Help](#)

Hint: Use Ctrl + Click for multiple selections, or Shift + Click for a range.

Gender
☐ All Gender
☐ Female
☐ Male

Race
☐ All Races
☐ American Indian or Alaska Native
☐ Asian or Pacific Islander
☐ Black or African American
☐ White

Hispanic Origin
☐ All Origins
☐ Hispanic or Latino
☐ Not-Hispanic or Latino
☐ Not Stated

Age
☐ All Ages
☐ <1 year
☐ 1-4 years
☐ 5-14 years
☐ 15-24 years
☐ 25-34 years
☐ 35-44 years
☐ 45-54 years
☐ 55-64 years
☐ 65-74 years
☐ 75-84 years
☐ 85+ years

Default rates per 100,000

4. Select year and month

[Send](#)
[Help](#)

Browse or search to find items in the Year/Month Finder Tool, then highlight the items to use for this request.
 (The Currently selected box displays all current request items.)

[Finder Tool Help](#) [Advanced Finder Options](#)

[Browse](#)
[Search](#)
[Delete](#)

Year/Month

☐ 2011 (2011)
☐ 2012 (2012)
☐ 2013 (2013)
☐ 2014 (2014)
☐ 2015 (2015)
☐ 2016 (2016)
☐ 2017 (2017)
☐ 2018 (2018)
☐ 2019 (2019)

Currently selected:
☐ All
☐ All Dates

Browse the list by opening and closing items.
 Use Ctrl+Click to multiply select, Shift+Click for a range.

5. Select weekday, autopsy and place of death

[Send](#)
[Help](#)

3

Load Data