**Project Title: Gun Deaths in America, An Analysis**

**Team Members:**

|  |
| --- |
| **Name** |
| Chris Janssen |
| Bhumika Sharma |
| Cong Tran |
| Vijaybabu Gangaprasad |
| Imran Bawany |

**Logo: Coming Soon!**

**Trello Board Screen Shot**:

*Screenshot of your workflow, data design, or Project Management Board that shows breakdown of tasks*

1. Explore Data: get familiar with all of the features, especially the demographic fields and the fields the fields that have to do with date, cause of death, and the ICD 10 codes. Use pd.describe() for the features of interest. Also do value\_counts().
2. Transformations: Transform features from the coded data dictionary into the

**Project Description(High Level):**

**Motivation (WHY you feel it’s valuable):**

**User Stories:**

*As a concerned citizen I want to understand the nature of gun violence in America so that we can have a safer and healthier society.*

**Libraries to be Used:**

Pandas

Matplotlib

Numpy

**Packages Required:**

**Goals:**

MVP Goals:

Analyze data for each hypothesis as they follow:

1. Does race play a factor in suicide by gun?
   1. Person: Chris
2. Does race play a factor in homicide by gun?
   1. Person: Vijay
3. Does gender play a role in suicide/homicide by gun?
   1. Person: Bhumika
4. Does race play a factor in legal intervention (i.e. police shootings) by gun?
   1. Person: Imran
5. Are gun deaths getting worse and by what type of death (suicide, homicide, or accident)
   1. Person: Cong (plot gun deaths by year 2015-2017 and by each category)
6. Have gun deaths by assault weapons/rifles/large weapons increased?
   1. Person: Chris
7. Do gun deaths get worse on the weekends?
   1. Person: Vijay

Stretch Goals:

1. Does the time of year play a factor in gun deaths, and by which types?
   1. Person: Bhumika
2. Does age play a role in gun deaths?
   1. Person: Cong
3. Can we create an “avatar” for the average gun death victim
4. Show the statistical significance for disparities like race, gender, etc.

**Breakdown of Tasks**(Ownership by Group Member)**:**

1. **Data Sanitization:** Bhumika will create a dataframe based on 2015, 2016, and 2017 gun deaths with the features we need. (Drop features that have bad data or too much missing data, or other unwanted columns)
2. **Data Analysis by feature category:** Two goals per team member.
3. **Statistics:** Run t-tests for each goal/trend or other relevant tests. Show standard error, median, mean, z-score. Do a logistic regression (calculate an odds ratio).
4. **Visualization:** Plot data (two different charts) for each goal/trend.
5. **Aggregation:** Put all of the charts into one presentation and rehearse it.
6. **Create logo:** Make the project look nice ☺

**Schedule for Completion of Tasks:**

|  |  |  |
| --- | --- | --- |
| **Date** | **Task** | **Notes** |
| **09/31** | Get data sanitized into one frame | By: Bhumika |
| **10/03** | Finish data analysis |  |
| **10/05** | Finish statistics |  |
| **10/07** | Project Work & Mock Presentations |  |
|  |  |  |
| **10/09** | PROJECT PRESENTATION |  |

**Models & Columns:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Model Name** | **Model Attributes** | **Model validations** | **Model Associations** |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

**Associations(Has Many/Belongs To):**

See the above goals since each of these hypothesis requires an association when being visualized/plotted.

**Git Workflow:**

**“Stuck Time” Agreement:**

* If somebody is not able to finish, we will have a group Webex discussion and figure out if there is a problem. If a member is stuck, another one could help. We aim to communicate clearly if there is a problem

**Retrospective Notes (Due DATE):**

**ALL Deployed & Repos Links (Due DATE):**

**Working Agreements:**

**Project Requirements:**

You must use at least two of the below:

* Use Pandas to clean and format your dataset(s).
* Create a Jupyter Notebook describing the **data exploration and cleanup** process.
* Create a Jupyter Notebook illustrating the **final data analysis.**
* Use Matplotlib to create a total of 6–8 visualizations of your data (ideally, at least 2 per “question” you ask of your data).
* Save PNG images of your visualizations to distribute to the class and instructional team, and for inclusion in your presentation.
* (Optional) Use at least one API, if you can find an API with data pertinent to your primary research questions.
* Create a write-up summarizing your major findings. This should include a heading for each “question” you asked of your data and a short description of your findings and any relevant plots.

**Expectations:**

* Prepare a 15-minute data deep-dive or infrastructure walkthrough that shows what we’ve already learned.
* We expect you to put serious time and thought into this.
* We expect you to report problems you are facing along the way.
* We expect you to utilize some form of project management system.
* We expect you to dig deep into documentation and external resources to learn what you need.

**Example projects:**

* **Private Investigator:** Use aggregate crime data from different police precincts in a city to uncover patterns in
* **Uber Rides and Weather**: find out if people take Uber more during summer and winter, and if there are relationships between daily temperature and ride frequency.
* **Bullying and Crime Rates**: Bullying and violent crime seem like they should be related. Can we find a correlation between frequency of bullying and   
  rates of violent crime?

**Notes:**