**ABSTRUCT:**

**About dataset:**

The **Fatal Accidents 2007** dataset contains all Fatal Accidents on public roads reported to the National Highway Transportation Safety Administration (NHTSA).

**Research questions:**

1. Fatalities by Crash Month, Day of Week, Hours, state
2. Crash Counts by Roadway Function Class
3. Fatal Crashes by Roadway Function Class and weather Condition
4. Fatal crashes occurred by drunk and drive and over speed.
5. crash counts by relation to junction and manner of collision and road way alignment
6. crash counts by route, surface condition, construction or maintenance zone
7. hit and run involved in crashes
8. school bus related vehicles involved in crash

**Techniques and tools:**

I want to use R and Hadoop.

descriptive and exploratory data analysis.

Exploratory Charts

Regression model

**Data set:**

****