



Reducing Traffic Mortality in USA

Presented by: Anita Durg
Last Updated: Jan 20, 2023



Objective

- Derive a strategy to reduce the incidence of road accidents across the nation.
- Find patterns in the variations between the US States to derive suggestions for a policy action plan.
- Find similar profile US States in a statistically sound way



Data Science Pipeline

- Data wrangling
- Plotting
- Dimensionality reduction
- Unsupervised clustering



Data Source

- <https://github.com/fivethirtyeight/data/tree/master/bad-drivers>
- <https://creativecommons.org/licenses/by/4.0/>
- <https://www.kaggle.com/>



Glimpse of the Dataset - accidents.csv

	state	drv_r_fatl_col_b miles	perc_fatl_speed	perc_fatl_alcohol	perc_fatl_1st_time
46	Virginia	12.7	19	27	88
47	Washington	10.6	42	33	86
48	West Virginia	23.8	34	28	87
49	Wisconsin	13.8	36	33	84
50	Wyoming	17.4	42	32	90



Dataset Column Abbreviations

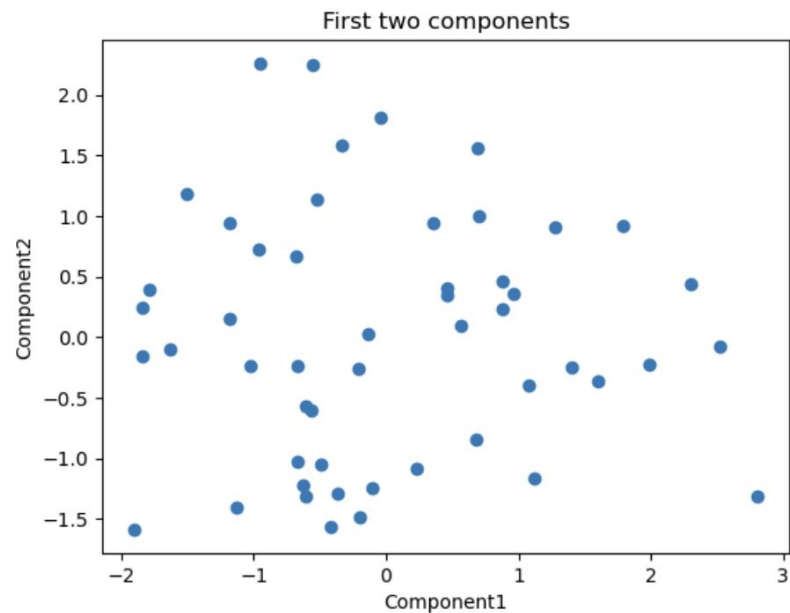
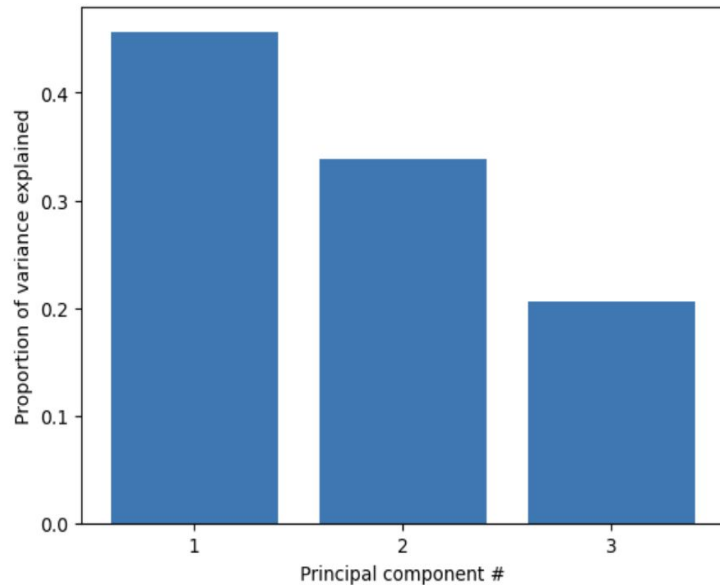
- `drvr_fatl_col_bmiles` = Number of drivers involved in fatal collisions per billion miles (2011)
- `perc_fatl_speed` = Percentage Of Drivers Involved In Fatal Collisions Who Were Speeding (2009)
- `perc_fatl_alcohol` = Percentage Of Drivers Involved In Fatal Collisions Who Were Alcohol-Impaired (2011)
- `perc_fatl_1st_time` = Percentage Of Drivers Involved In Fatal Collisions Who Had Not Been Involved In Any Previous Accidents (2011)



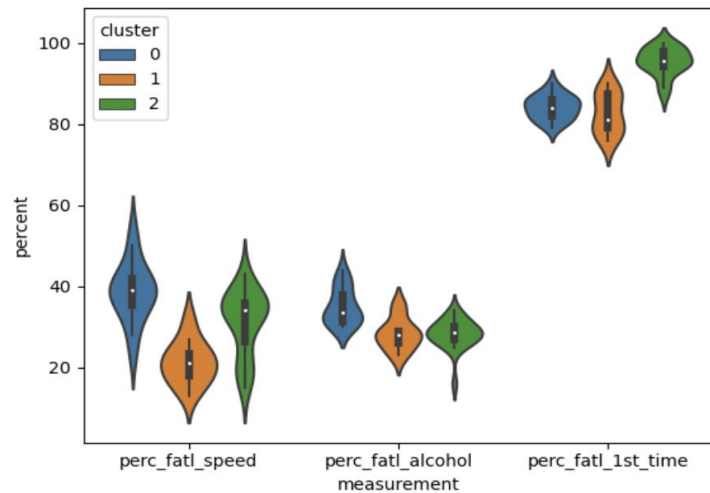
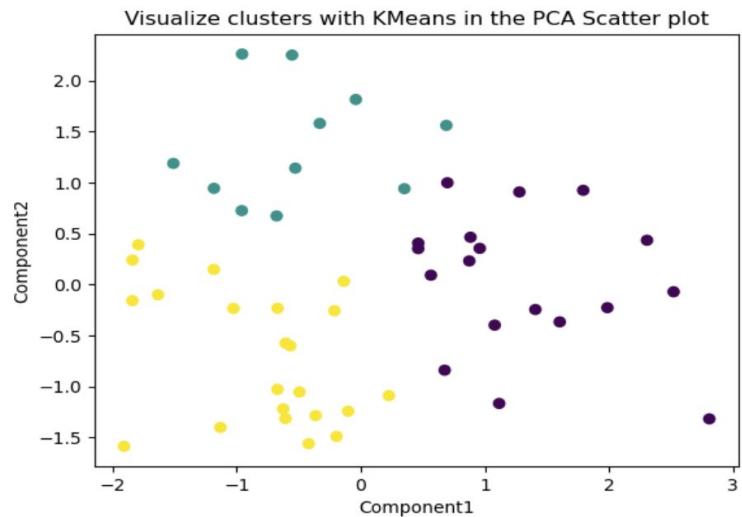
Association of features and accidents

	drvr_fatl_col_bmiles	perc_fatl_speed	perc_fatl_alcohol	perc_fatl_1st_time
drvr_fatl_col_bmiles	1.000000	-0.029080	0.199426	-0.017942
perc_fatl_speed	-0.029080	1.000000	0.286244	0.014066
perc_fatl_alcohol	0.199426	0.286244	1.000000	-0.245455
perc_fatl_1st_time	-0.017942	0.014066	-0.245455	1.000000

Perform PCA

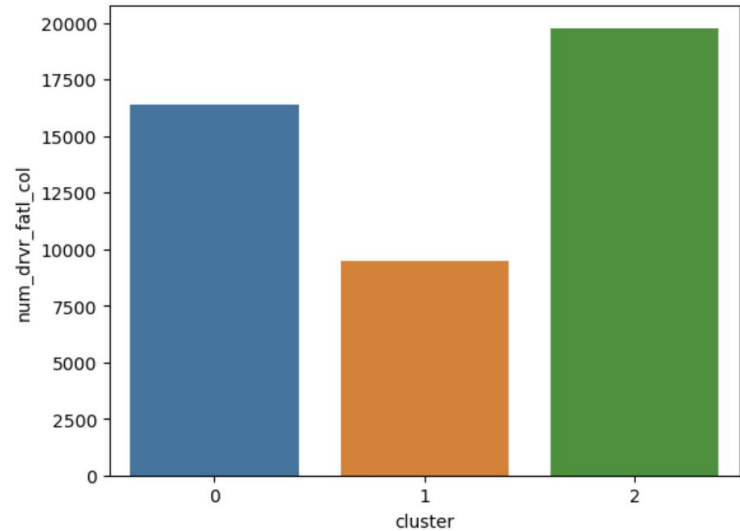


Visualize Clusters & visualize feature differences between the clusters



Number of accidents within each cluster

	count	mean	sum
cluster			
0	18	911.406439	16405.3159
1	11	860.505945	9465.5654
2	22	898.378595	19764.3291





Conclusions

- US States belonging to Cluster 0 should be a focus for policy intervention and further investigation
- The mean for number of accidents is highest for cluster 0 and therefore those states need to take stringent measures on alcohol policy and policy for maintaining speed limits



Acknowledgements

- Datacamp
- Kaggle
- Springboard Data Science Career Track
- My mentor A J Sanchez