INFO CHALLENGE 2023

TEAM IC23006

Members:

- Manas M Bhat
- Shantanu Parab
- Vineet Singh

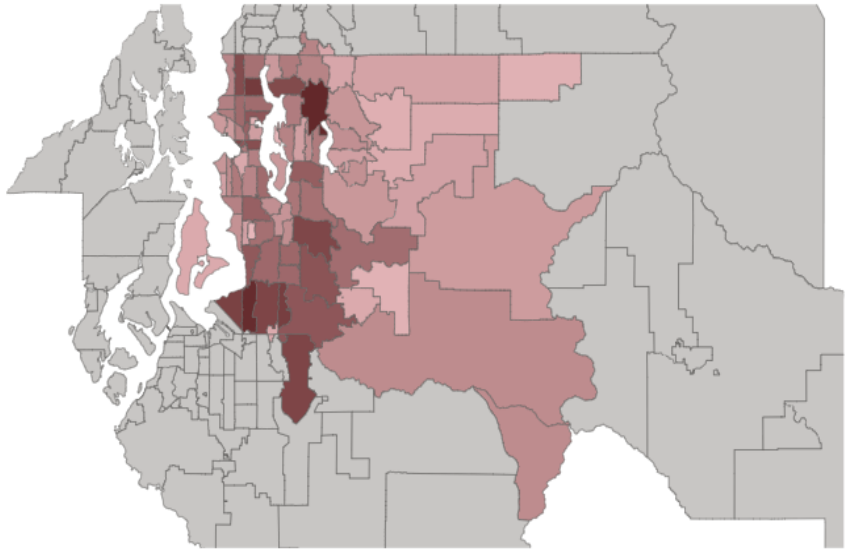
ANALYZING THE DATASET

- The dataset is a collection of survey from KING County for the years 2018, 2019, 2021, 2022.
- The year wise distribution of data is shown

DIVERSITY OF DATA

- Segregation of data based on Gender, Ethnicity and Education.

Responses



Races

45	124	63	2493	35	440	441	202
Sum of American...	Sum of Multiracial	Sum of Other Race	Sum of White	Sum of Native Ha...	Sum of Hispanic	Sum of Asian	Sum of Black

Gender

1936	1856	51
Sum of Female	Sum of Male	Sum of Other Gender

Education

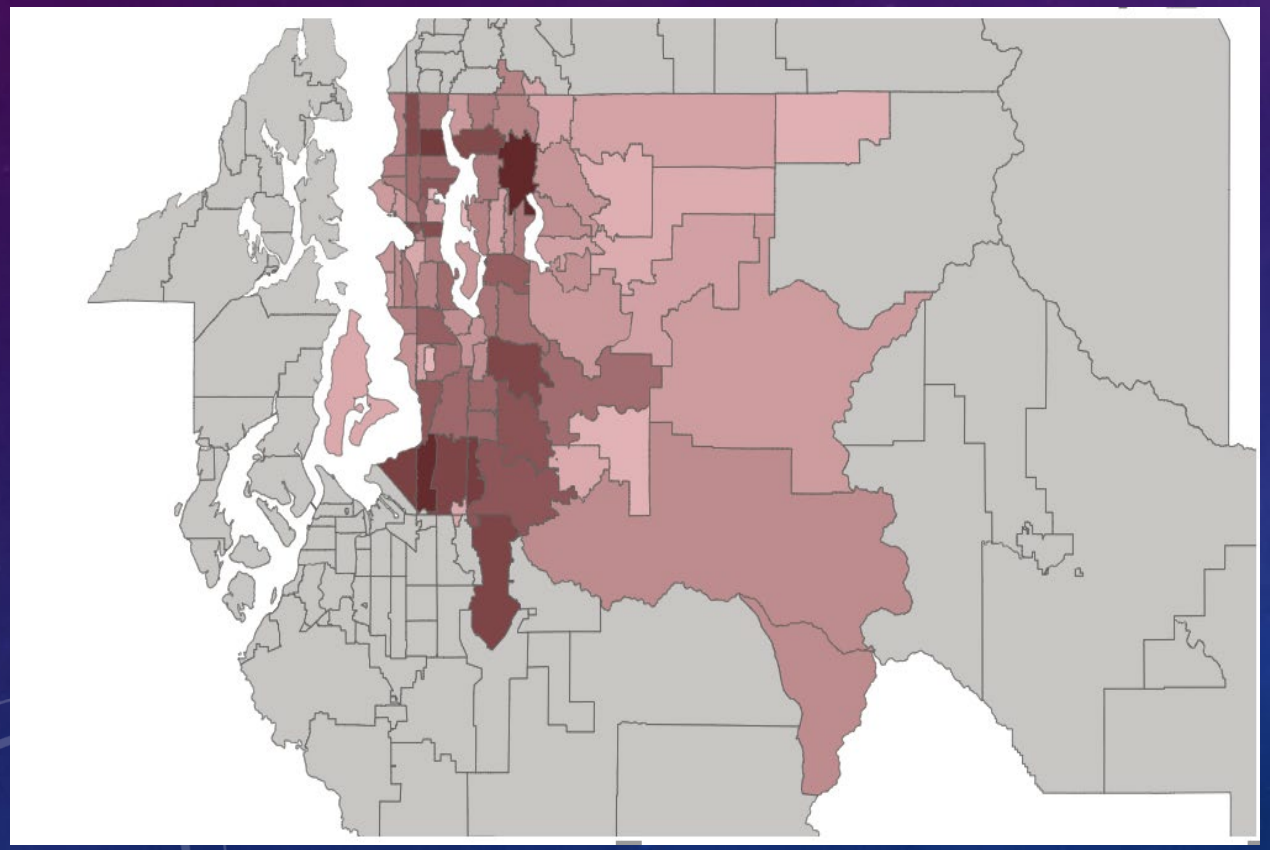
1492	759	678	914
Sum of Associate/Bachelors	Sum of High School	Sum of Masters Or Higher	Sum of Some College

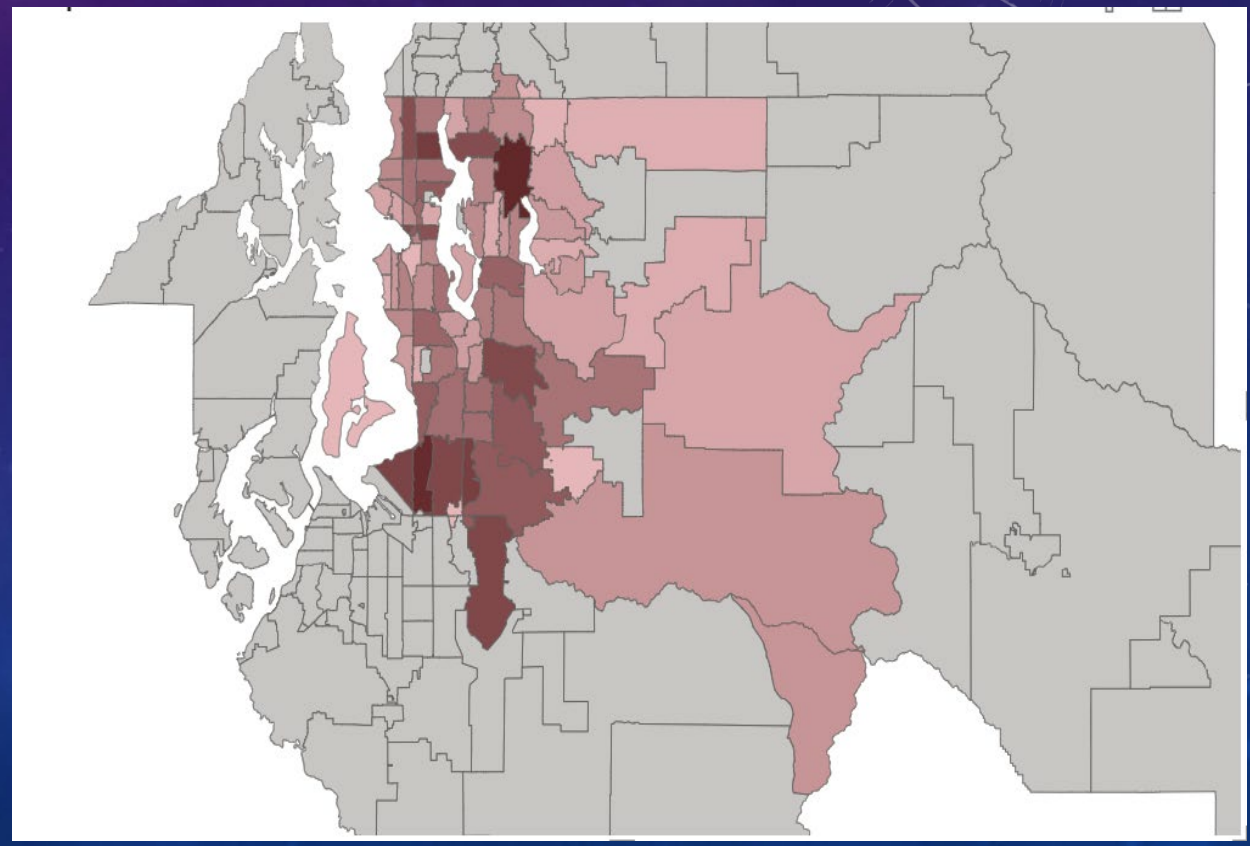
Income

1088	413	481	1117	744
Sum of < $40K	Sum of > $160K	Sum of $120 - $160K	Sum of $40K - $80K	Sum of $80K - $120K

REMOVING OUTLIERS

- Minimum 10 participants to be present for the zip code to be considered for the analysis.



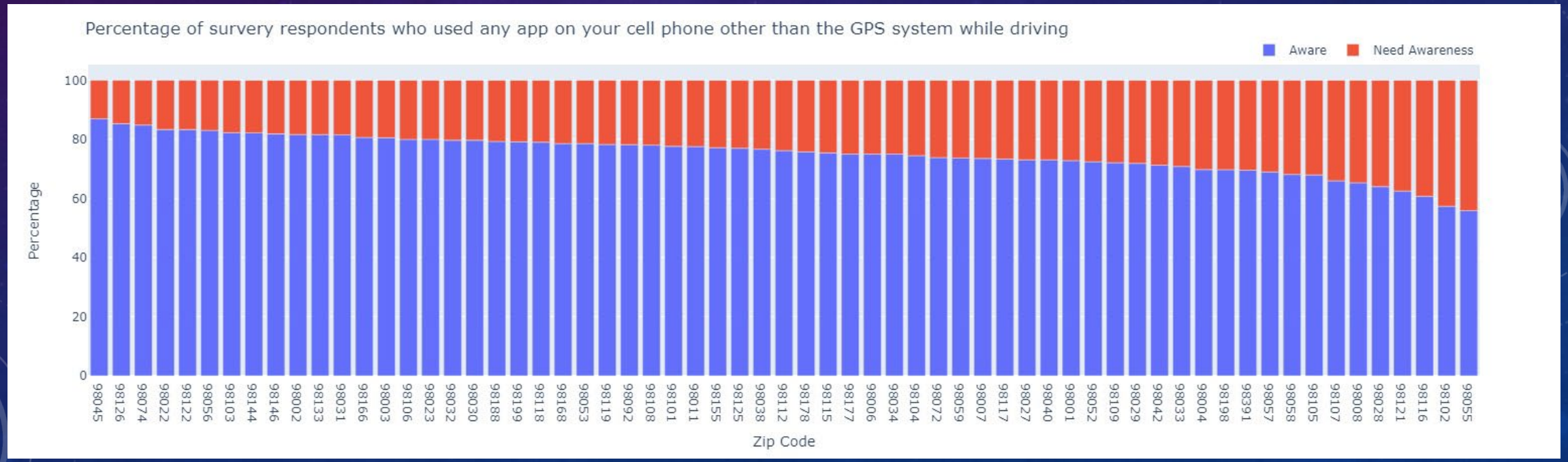


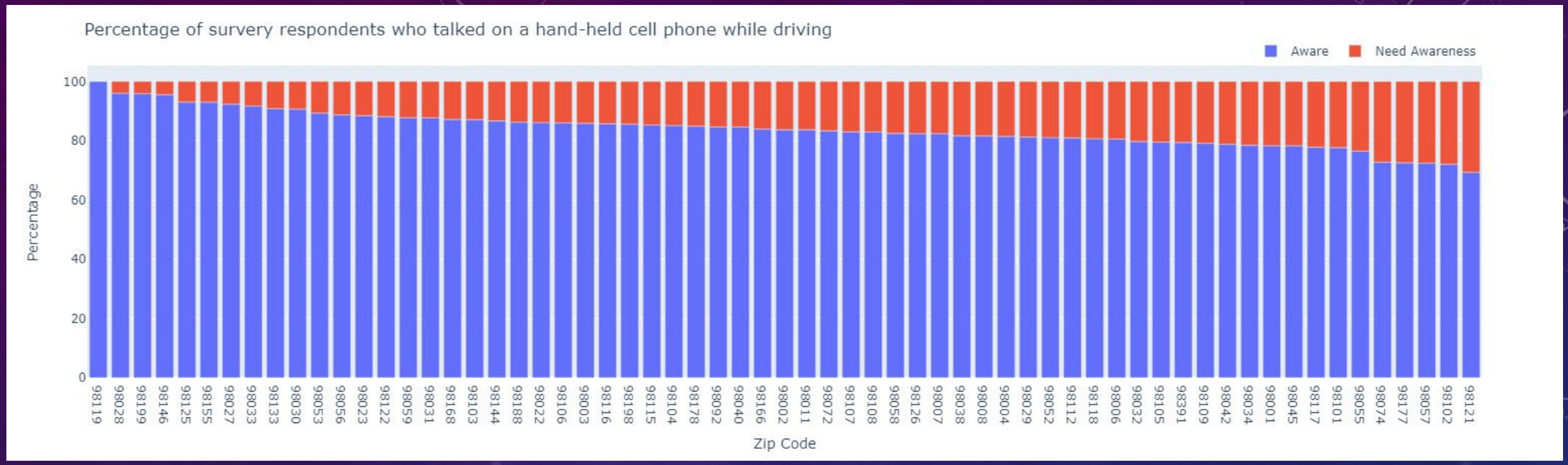
PROBLEM STATEMENTS

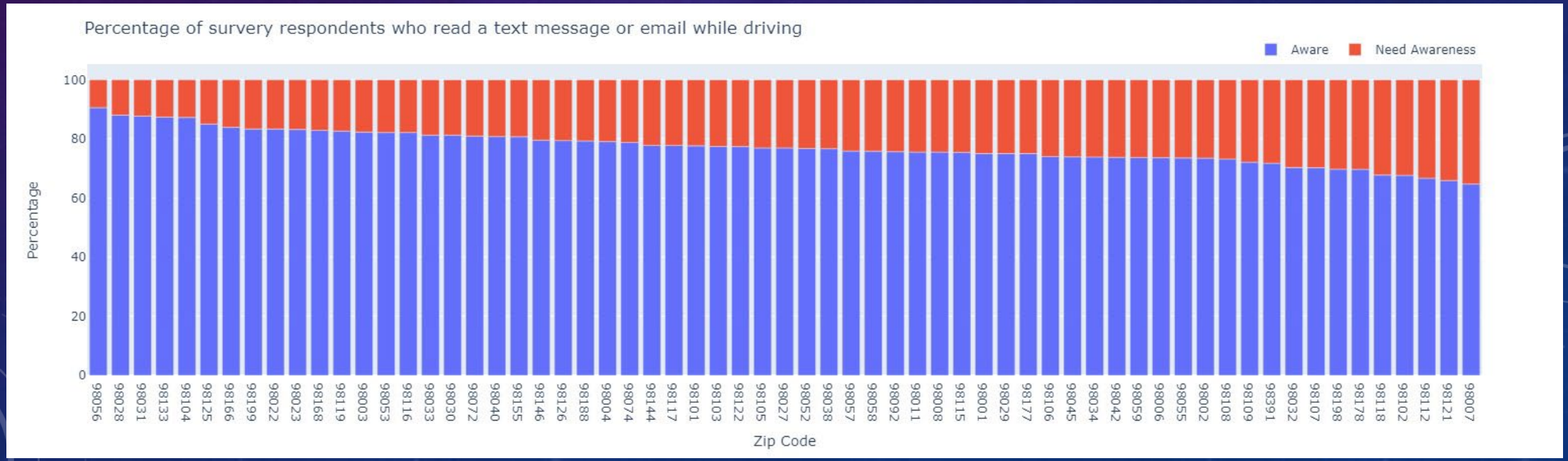
- To identify the key behavioral problems across different zip codes
- To generate a score for each respondent to analyze the general awareness pattern.
- To identify some important methods to curb dangerous driving behaviors.
- To understand the reliability of the data.

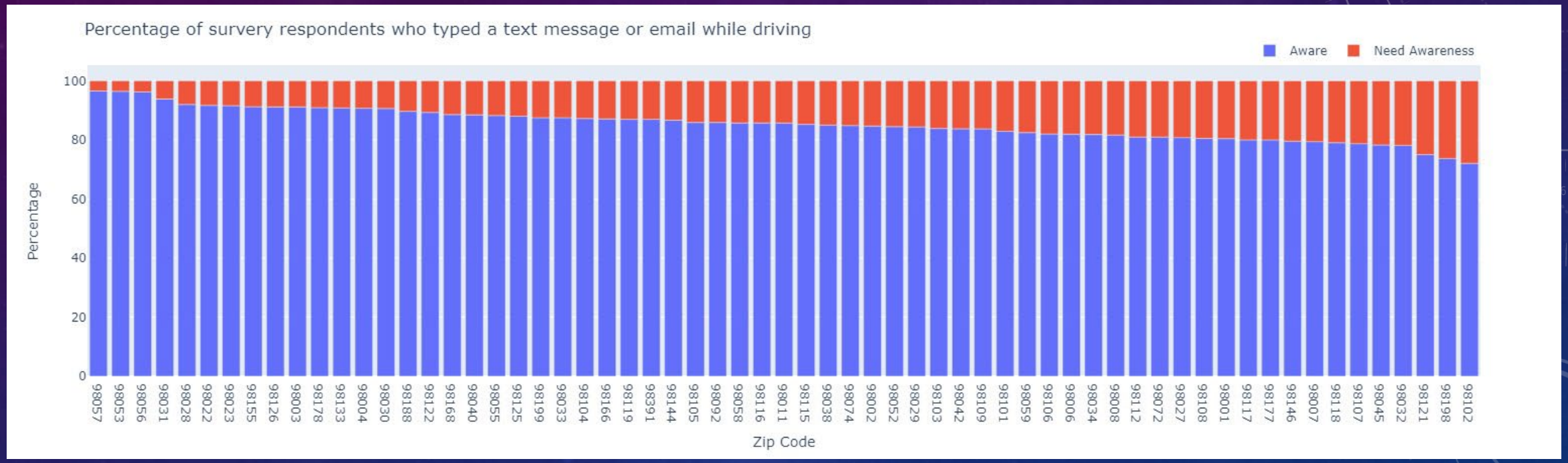
DANGEROUS BEHAVIOR ANALYSIS

- Data is divided for 4 dangerous behaviors.
  - Typing while driving
  - Reading while driving
  - Talking on cellphone while driving
  - Using apps while driving (excluding GPS)




Percentage of survery respondents who talked on a hand-held cell phone while driving
Aware
Need Awareness
Percentage
0
20
40
60
80
100
98119 98028 98199 98146 98125 98155 98027 98033 98133 98030 98053 98056 98023 98122 98059 98031 98168 98103 98144 98188 98022 98106 98003 98116 98198 98115 98104 98178 98092 98040 98166 98002 98011 98072 98107 98108 98058 98126 98007 98038 98008 98004 98029 98052 98112 98118 98006 98032 98105 98391 98109 98042 98034 98001 98045 98117 98101 98055 98074 98177 98057 98102 98121
Zip Code


Percentage of survery respondents who read a text message or email while driving
Aware
Need Awareness
Percentage
0
20
40
60
80
100
98056 98028 98031 98133 98104 98125 98166 98199 98022 98023 98168 98119 98003 98053 98116 98033 98030 98072 98040 98155 98146 98126 98188 98004 98074 98144 98117 98101 98103 98122 98105 98027 98052 98038 98057 98058 98092 98011 98008 98115 98001 98029 98177 98106 98045 98034 98042 98059 98006 98055 98002 98108 98109 98391 98032 98107 98198 98178 98118 98102 98112 98121 98007
Zip Code

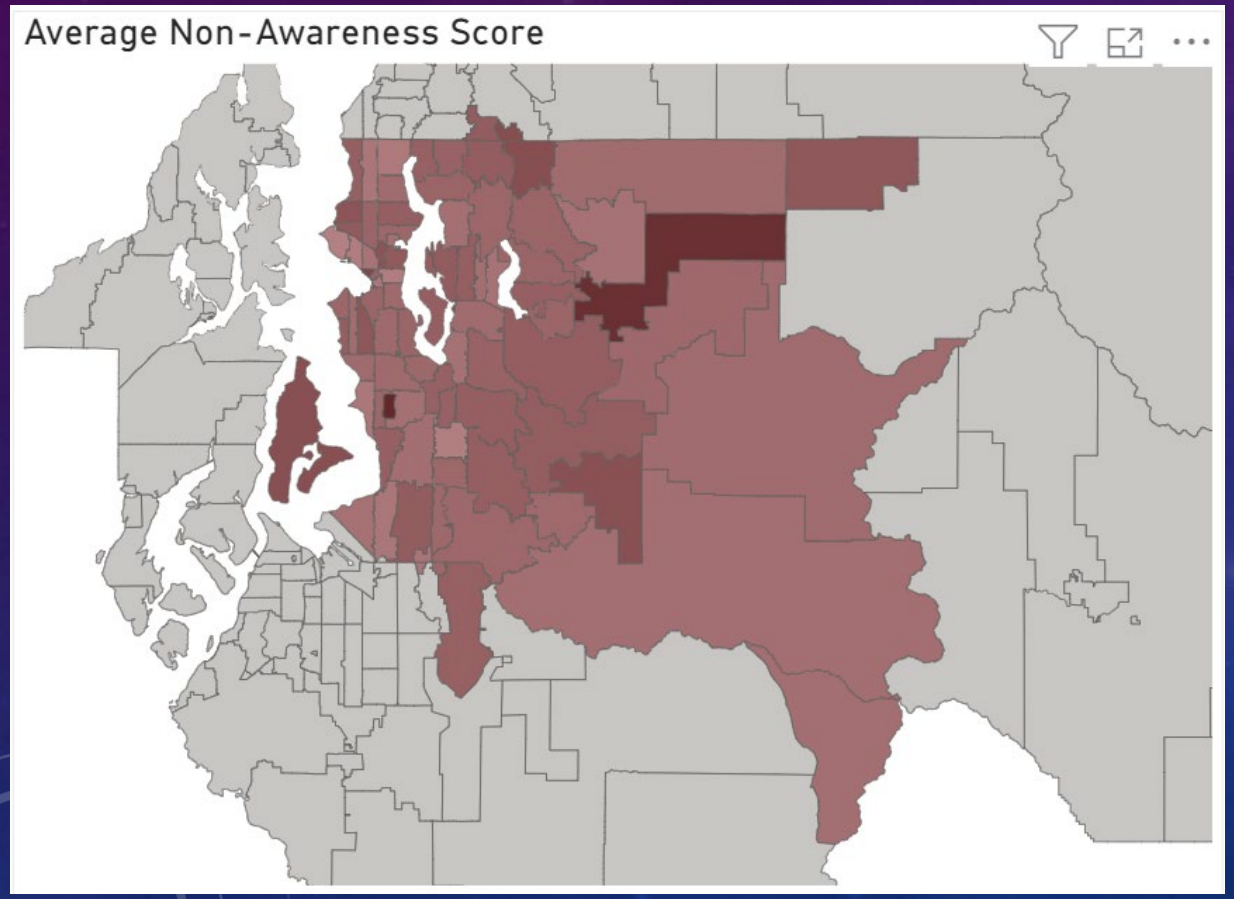

Percentage of survery respondents who typed a text message or email while driving
Aware
Need Awareness
Percentage
100
80
60
40
20
0
98057
98053
98056
98031
98028
98022
98023
98155
98126
98003
98178
98133
98004
98030
98188
98122
98168
98040
98055
98125
98199
98033
98104
98166
98119
98391
98144
98105
98092
98058
98116
98011
98115
98038
98074
98002
98052
98029
98103
98042
98109
98101
98059
98106
98006
98034
98008
98112
98072
98027
98108
98001
98117
98177
98146
98007
98118
98107
98045
98032
98121
98198
98102
Zip Code

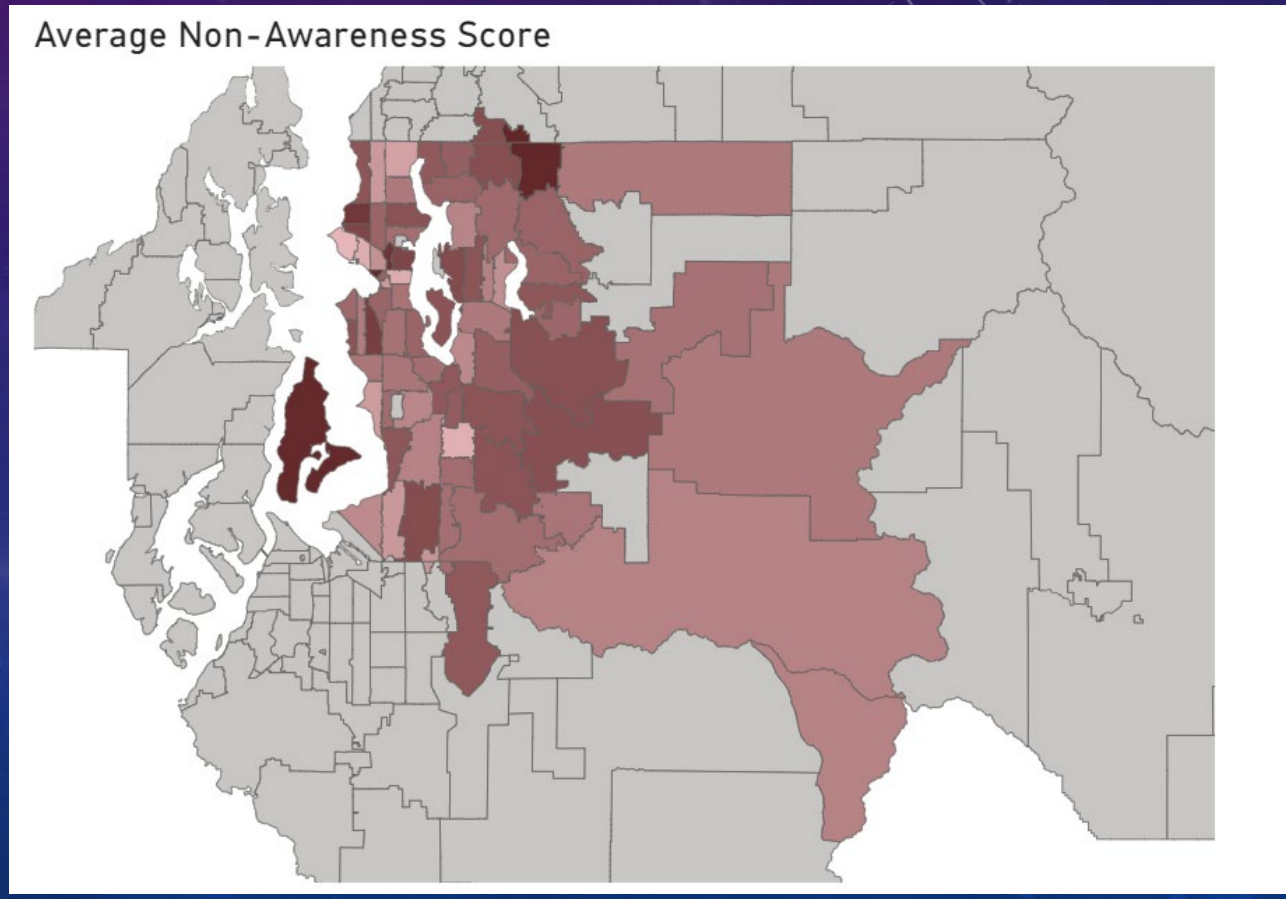
ASSIGNING A QUANTITATIVE SCORE

- We devised a scoring method based on severity of behaviors and opinions of individuals.
- Different weights are assigned for different parameters.
- **Nonawareness score** = Behavior score*2 + opinion on other driver's behavior *1 + how much of threat an activity is*1 + not knowing legal behavior*3 + How likely they are to commit illegal behavior*3 +Not a regular driver*0.5

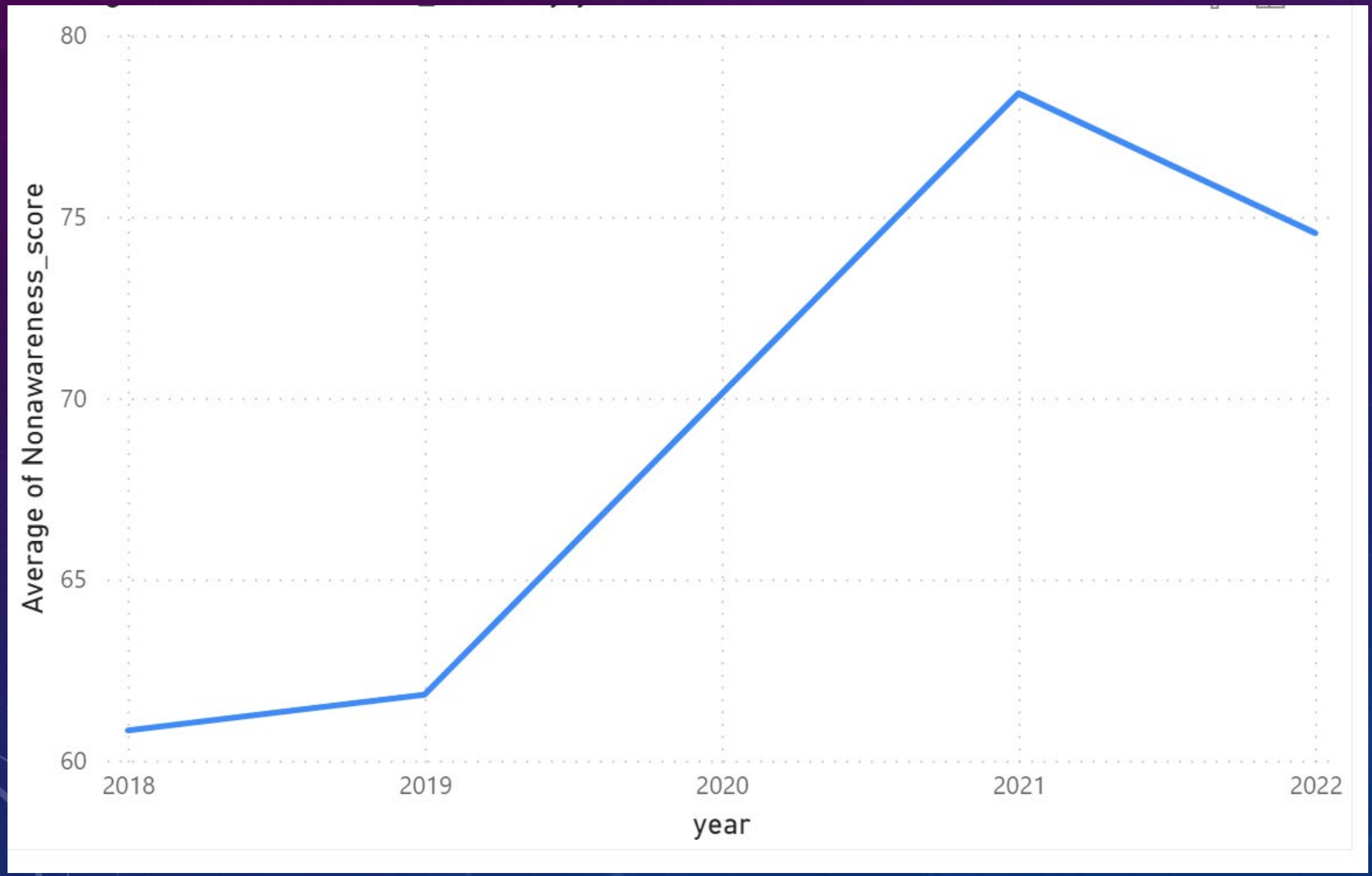
RESULTS FOR QUANTITATIVE SCORE

- Need for awareness graph for each zip code as calculated from the Nonawareness score.



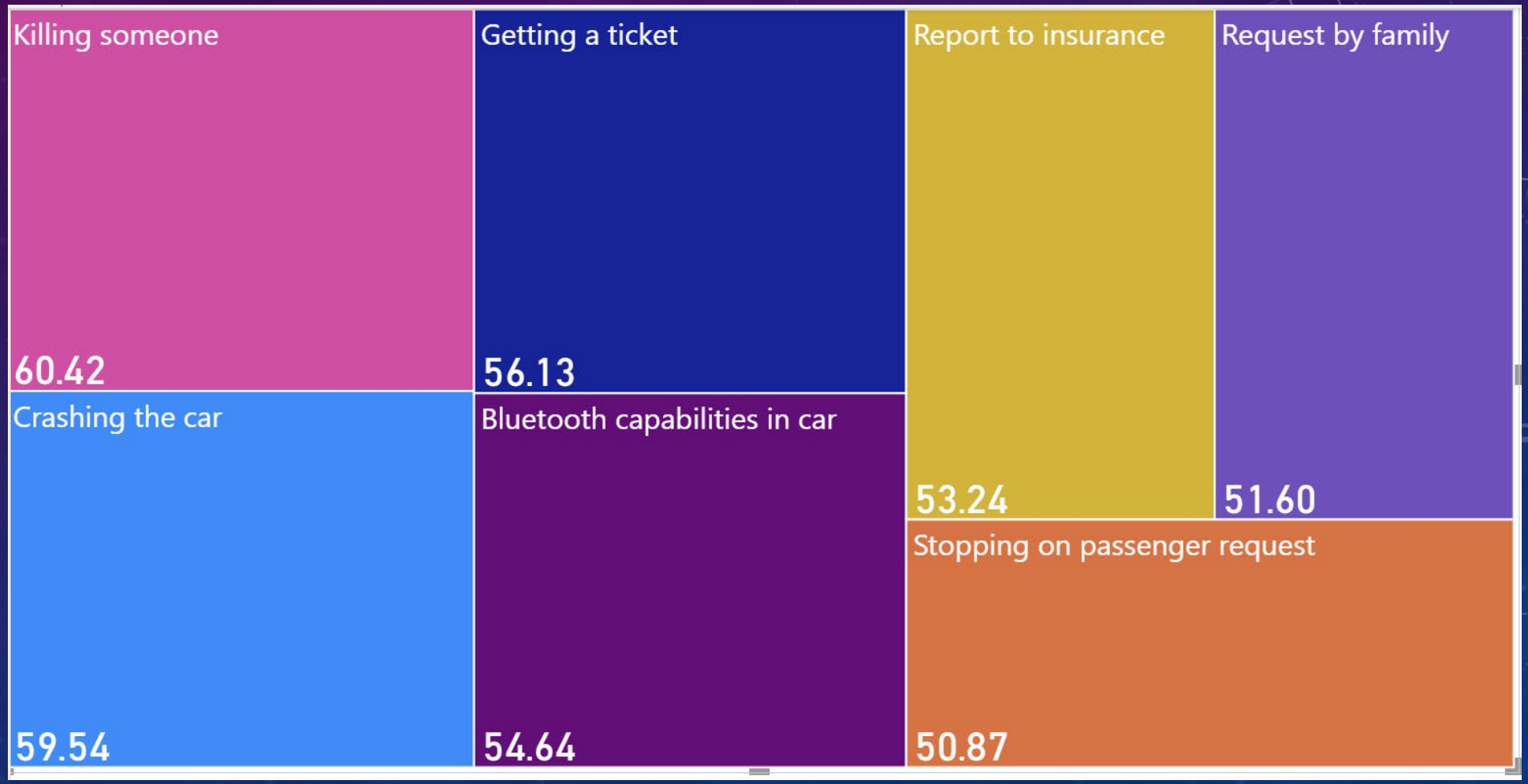


NON-AWARENESS CHANGING YEAR ON YEAR



SPREADING AWARENESS

- Best ways of spreading awareness using most receptive methods as suggested by participants themselves.



IS THE DATA RELIABLE ????

```
traffic_data[(traffic_data['drive30days']==2) & ((traffic_data['beh_read']>=4) | (traffic_data['beh_type']>=4)|
                                                 (traffic_data['beh_handheld']>=4)| (traffic_data['beh_handfree']>=4)| (traffic_data['beh_app']>=4))]
```

[91]

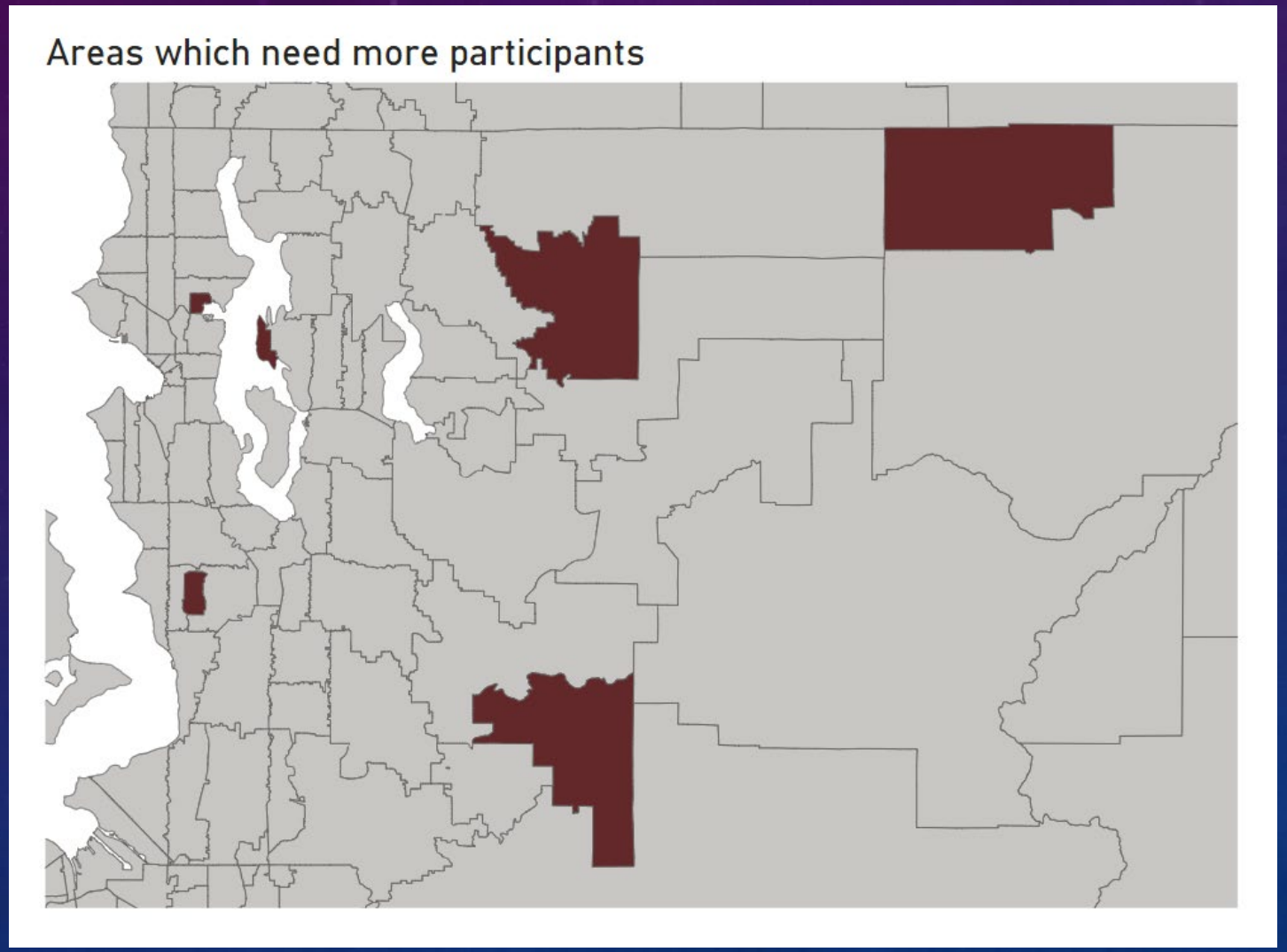
Python

	zip	work_outside_home	Commute	beh_read	beh_type	beh_handheld	beh_handfree	beh_app	acc_handheld	acc_handfree	acc_read	acc_type	acc_app	threat_talk	threat_type	threat_aggressive	threat_s
1	98052	1	1.0	5	5	5	2	3	4	4	4	4	4	2	1	1	
3	98028	2	NaN	5	5	5	5	5	4	1	4	4	4	1	1	2	
5	98006	2	NaN	5	5	5	3	5	3	1	3	3	3	2	1	1	
6	98109	1	3.0	5	5	5	5	5	4	3	3	4	4	1	1	1	
12	98122	1	3.0	5	5	5	5	5	4	2	4	4	4	2	2	2	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	
3822	98008	2	NaN	5	5	5	5	5	4	3	4	4	4	1	1	1	
3832	98002	1	2.0	1	5	5	4	5	2	1	3	4	4	3	1	1	
3834	98112	2	NaN	5	5	5	5	5	3	2	4	4	3	3	1	1	
3836	98122	1	3.0	6	6	6	5	5	3	2	3	4	4	2	2	1	
3838	98122	2	NaN	5	5	5	5	5	4	2	4	4	4	1	1	1	

653 rows × 72 columns

NEED MORE SURVEY PARTICIPANTS

- Areas where participation was not sufficient needs more awareness.



THANK YOU