Build model and score on as feature based on accident data:

Accident Type (ie prob of each accident type)

Similar for:

* "HUMAN\_COST"
* "TOTAL\_COST"
* ROAD\_GEOM

Importance for Total Cost (excl lat long)

Feature Gain Cover Frequency

1: DAY\_OF\_WEEK 0.23957625 0.17114140 0.13069307

2: QUARTER 0.17974393 0.17149712 0.25698570

3: ROAD\_TYPE 0.16635146 0.13063781 0.08690869

4: URBAN\_TYPE 0.09513233 0.14529313 0.11958196

5: LIGHT 0.08559235 0.06774620 0.12233223

6: ROAD\_GEOM 0.07021821 0.07454741 0.07832783

7: LGA\_NAME 0.06322485 0.08752388 0.08294829

8: ROAD\_INT\_TYPE 0.05905259 0.08303274 0.07051705

9: SPEED\_ZONE 0.04110803 0.06858032 0.05170517

Importance for Human Cost – ie similar need to build similar features

Feature Gain Cover Frequency

1: DAY\_OF\_WEEK 0.23626220 0.17189856 0.13219909

2: QUARTER 0.17902329 0.19425095 0.25213991

3: ROAD\_TYPE 0.16655880 0.12749981 0.08760435

4: URBAN\_TYPE 0.08941488 0.14189139 0.11201522

5: LIGHT 0.08660742 0.07051263 0.12649266

6: ROAD\_GEOM 0.07266085 0.07152209 0.08147522

7: LGA\_NAME 0.07117646 0.08640761 0.08950650

8: ROAD\_INT\_TYPE 0.05768416 0.07218825 0.06731481

9: SPEED\_ZONE 0.04061194 0.06382872 0.05125225

If we include, number of vehciles (by category eg, CAR, BUS etc) then for TOTAL\_COST we get

Feature Gain Cover Frequency

1: CAR 0.20858589 0.27299544 0.21912480

2: NO\_PERSONS 0.18856983 0.21195422 0.20265802

3: DAY\_OF\_WEEK 0.13216389 0.08351459 0.07209076

4: ROAD\_TYPE 0.10100674 0.06256521 0.04810373

5: QUARTER 0.10083298 0.08806519 0.14178282

6: URBAN\_TYPE 0.05266205 0.06985619 0.06495948

7: LIGHT 0.04621397 0.03509209 0.06651540

8: NO\_VEHICLES 0.03951366 0.02551159 0.02437601

9: LGA\_NAME 0.03853250 0.04127340 0.05089141

10: ROAD\_GEOM 0.03761702 0.03964414 0.04194489

11: ROAD\_INT\_TYPE 0.03085425 0.03851851 0.03727715

12: SPEED\_ZONE 0.02344723 0.03100943 0.03027553

And for Human Cost we get

Feature Gain Cover Frequency

1: CAR 0.20012558 0.26702241 0.21455422

2: NO\_PERSONS 0.19096783 0.22268026 0.20476630

3: DAY\_OF\_WEEK 0.13683592 0.08474443 0.07284205

4: QUARTER 0.10422795 0.08595514 0.14703170

5: ROAD\_TYPE 0.09650404 0.06614994 0.04560607

6: URBAN\_TYPE 0.04919334 0.06407213 0.06397617

7: LIGHT 0.04653273 0.03318759 0.06993404

8: ROAD\_GEOM 0.04177417 0.03478621 0.04291084

9: NO\_VEHICLES 0.04070188 0.02770809 0.02248386

10: LGA\_NAME 0.03908526 0.04458327 0.04858501

11: ROAD\_INT\_TYPE 0.02855452 0.03661551 0.03546351

12: SPEED\_ZONE 0.02549676 0.03249503 0.03184623

**Take away = Need to model CAR and NO\_PERSONS and other top factors here.**