# Vehicle Reliability Ranking

Spring 2022: Data Science Final Project Presentation
by Ray Ogunjimi

# Introduction

- “Japanese cars are reliable”
- “Hondas and Toyotas run for hundreds of thousands of miles”
- “German cars are precise but expensive to repair:
- “Regular oil changes can extend the life of any vehicle”

# How do we get the data

- Compare based on
  - Parts prices
  - Repair center/mechanic database
  - Average mileage
  - Number and age cars taken to junk/scrap yard
- Data is very hard to find but mileage price and service history can help

# Carfax

- All cars run and drive
- Repair history is always available
- Can sort such that cars have no reported accidents
- Reliable information

# Scraping the web and making API Calls

- Done in python outputted to a CSV
- Web Scraping is the least elegant task a programmer ever has to do
- Reverse engineering
- Not knowing problems until they come up
- CAPTCHA
- Very time consuming
- VIN can be decoded using an API
- Nice thing is that the data is in the exact format I need

# Scraping the web and making API Calls

https://github.com/rayogunjimi/Reliability-Ranker/blob/main/get_vins.py

https://github.com/rayogunjimi/Reliability-Ranker/blob/main/vin_lookup.py

# Wrangling

- CSV is imported
- Show service possibilities
- Count service per category
    - Powertrain repair (engine, transmission), maintenance service (oil, brake, tire), suspension, cosmetic

# Wrangling

https://github.com/rayogunjimi/Reliability-Ranker/blob/main/reliability_ranker.Rmd

# Outcomes and Visualizing

- Most common vehicle
- Average price based on body type, based on make
- Most common service
- Most common service center/mechanic
- Average car prices per make model year and mileage
- Number of repairs per 1000 miles (exclude low number of samples)

# Most Popular Services

| services<br><chr> | n<br><int> |
|---|---|
| Oil and filter changed | 19541 |
| Maintenance inspection completed | 18154 |
| Pre-delivery inspection completed | 7881 |
| Tires rotated | 7490 |
| Emissions or safety inspection performed | 6207 |
| Tire condition and pressure checked | 6086 |
| Vehicle washed/detailed | 5877 |
| Emissions inspection performed | 5015 |
| Brakes checked | 3828 |
| Fluids checked | 3766 |

# Most Popular Car Makes

| | |
|---|---|
| CHEVROLET | 557 |
| AUDI | 496 |
| NISSAN | 392 |
| TOYOTA | 352 |
| HONDA | 346 |
| BMW | 344 |
| KIA | 344 |
| MERCEDES-BENZ | 341 |
| FORD | 334 |
| LEXUS | 236 |
| HYUNDAI | 231 |
| ACURA | 221 |
| SUBARU | 205 |
| JEEP | 200 |
| LINCOLN | 192 |

# Most Popular Car Models

| | make <chr> | model <chr> | n <int> |
|---|---|---|---|
| 2 | SUBARU | Outback | 84 |
| 3 | MERCEDES-BENZ | GLC-Class | 81 |
| 4 | ACURA | RDX | 67 |
| 5 | TOYOTA | Highlander | 67 |
| 6 | CHEVROLET | Corvette | 64 |
| 7 | ALFA ROMEO | Giulia (952) | 62 |
| 8 | AUDI | Q5 | 62 |
| 9 | ACURA | TLX | 59 |
| 10 | AUDI | Q7 | 57 |
| 11 | MERCEDES-BENZ | C-Class | 56 |

# Most Popular Car Models and Years

| make <chr> | model <chr> | year <chr> |
|---|---|---|
| INFINITI | QX50 | 2019 |
| MERCEDES-BENZ | GLC-Class | 2019 |
| SUBARU | Outback | 2019 |
| LINCOLN | MKC | 2019 |
| ACURA | RDX | 2019 |
| SUBARU | Impreza | 2019 |
| TOYOTA | Highlander | 2019 |
| ALFA ROMEO | Giulia (952) | 2019 |
| AUDI | Q5 | 2019 |
| AUDI | Q8 | 2019 |
| ACURA | MDX | 2019 |
| ACURA | TLX | 2019 |
| LINCOLN | Aviator | 2020 |
| VOLVO | S60 | 2019 |
| ALFA ROMEO | Giulia (952) | 2018 |

# Makes with Most Mileage

| make <chr> | average_mileage <dbl> |
|---|---|
| SUZUKI | 110728.333 |
| MERCURY | 108113.800 |
| SAAB | 103454.500 |
| PONTIAC | 83774.000 |
| SATURN | 68774.000 |
| HUMMER | 65382.000 |
| JEEP | 63538.041 |
| TOYOTA | 62340.568 |
| RAM | 59630.244 |
| HONDA | 58748.940 |
| DODGE | 58220.673 |
| CHRYSLER | 57678.388 |
| HYUNDAI | 52846.878 |
| BUICK | 50483.260 |
| NISSAN | 49957.885 |

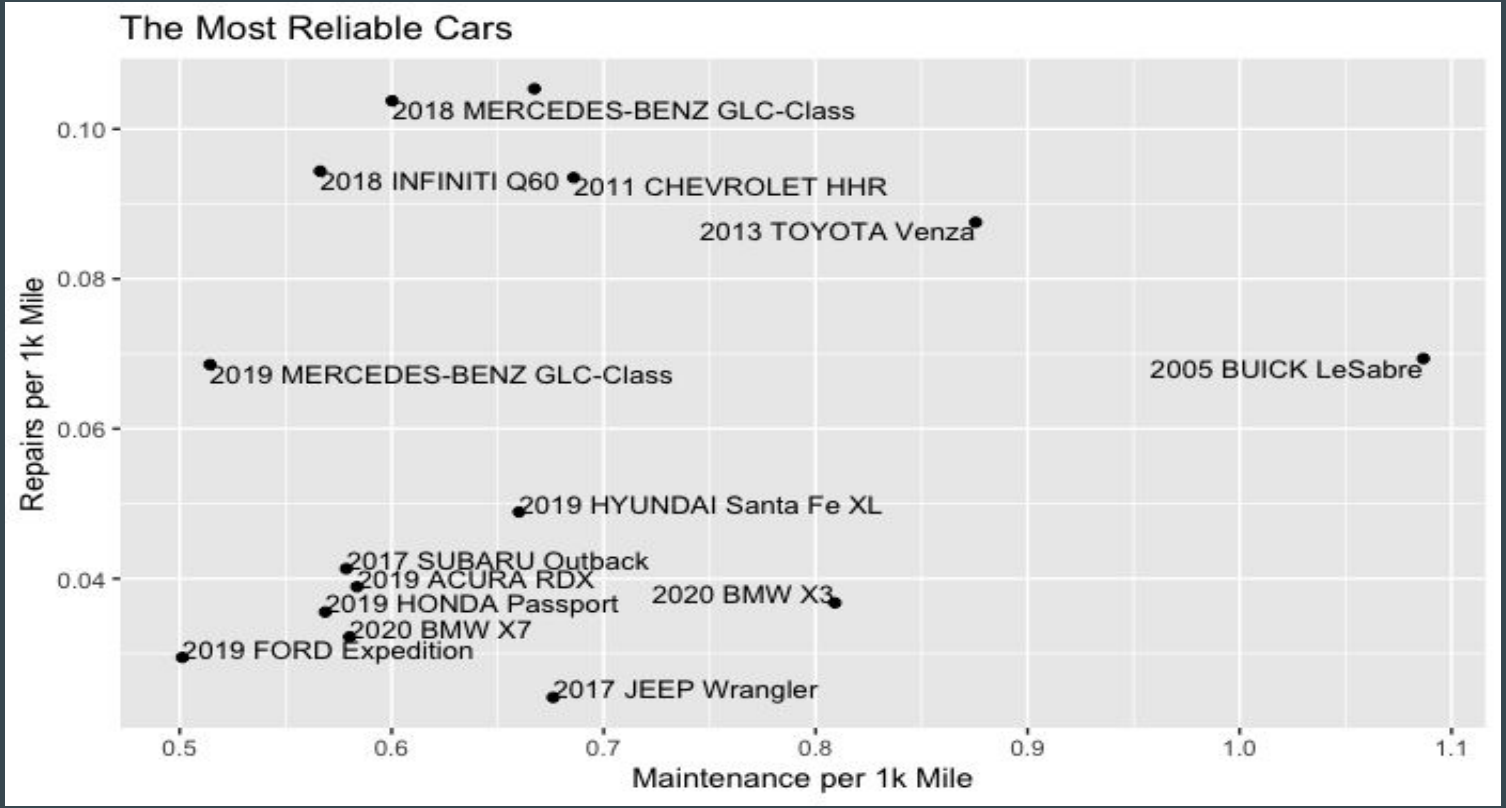
# Models with Most Mileage

| make <chr> | model <chr> | average_mileage <dbl> |
|---|---|---|
| MAZDA | Millenia | 214506.000 |
| DODGE | Dakota | 203726.000 |
| CHEVROLET | Astro Van | 186341.000 |
| TOYOTA | Echo | 185843.000 |
| INFINITI | FX35 | 174272.000 |
| GMC | Safari | 159613.000 |
| TOYOTA | Corolla Matrix | 157329.000 |
| MAZDA | Mazda2 | 155783.000 |
| FORD | Windstar | 152310.000 |
| MERCURY | Milan | 151929.000 |

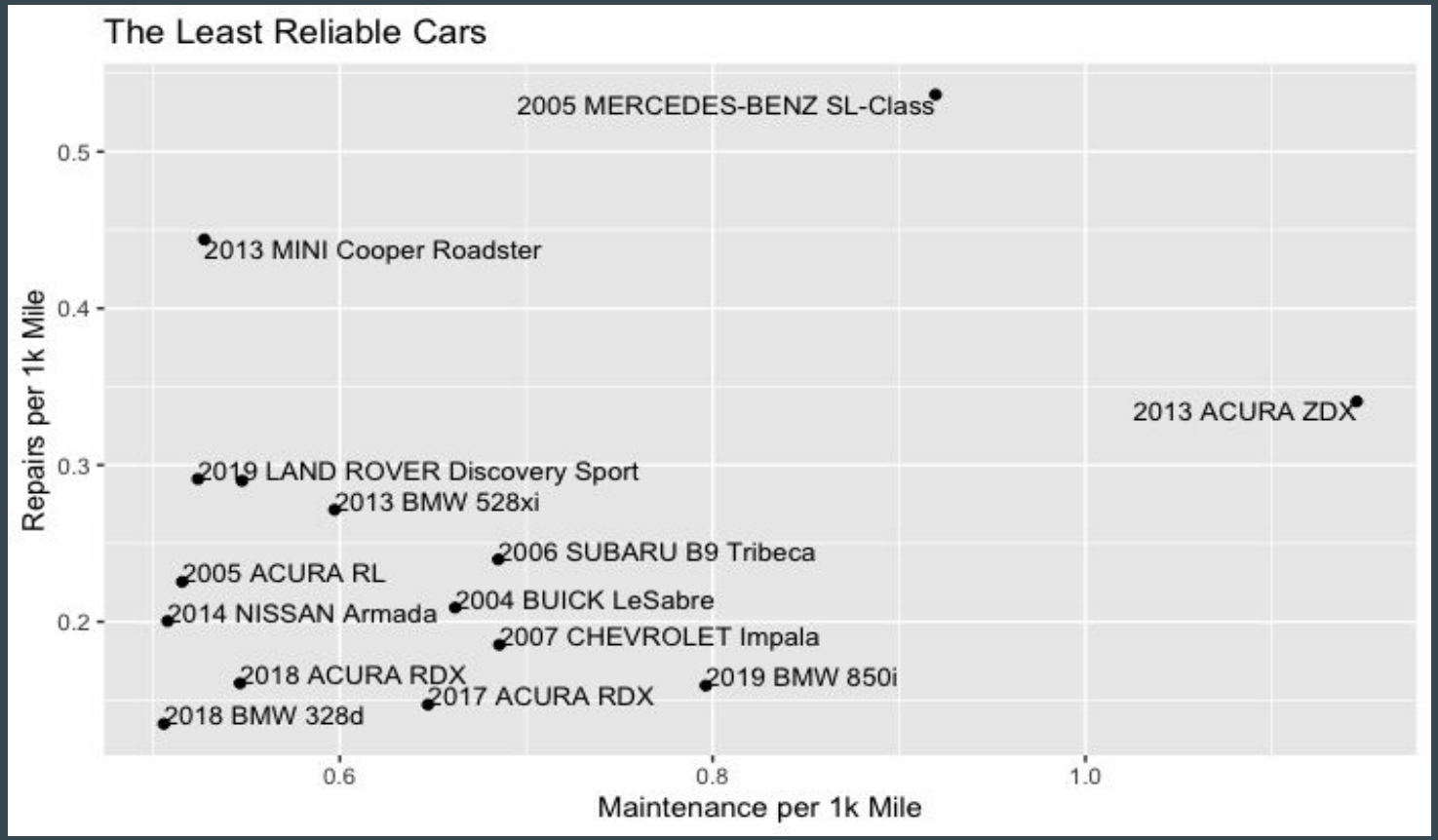
# Model Years with Most Mileage

| make <chr> | model <chr> | year <chr> | average_mileage <dbl> |
|---|---|---|---|
| TOYOTA | FJ Cruiser | 2008 | 277598.00 |
| LEXUS | GX | 2007 | 250952.00 |
| TOYOTA | Land Cruiser | 2000 | 247578.00 |
| NISSAN | Titan | 2015 | 246081.00 |
| NISSAN | NV200 | 2018 | 239622.00 |
| LEXUS | LX | 2006 | 224842.00 |
| HONDA | Ridgeline | 2011 | 220525.00 |
| MAZDA | Millenia | 2001 | 214506.00 |
| CHEVROLET | HHR | 2007 | 213695.00 |
| CHEVROLET | Trailblazer | 2006 | 210038.00 |

# Most Reliable



# Least Reliable



# Fast Depreciation

| make <chr> | model <chr> | year <chr> | coefficient <chr> | intercept <chr> | freq <int> |
|---|---|---|---|---|---|
| MERCEDES-BENZ | S-Class | 2019 | -1.75218557618863 | 142630.395472007 | 15 |
| MASERATI | Levante | 2019 | -0.592832870105253 | 81664.5232770117 | 12 |
| FORD | Transit | 2020 | -0.499125509560411 | 73978.6497535879 | 17 |
| LAND ROVER | Defender | 2020 | -0.486787749519915 | 90662.4145658607 | 12 |
| LAND ROVER | Range Rover Sport | 2019 | -0.419562200934997 | 90262.8818195689 | 11 |
| MASERATI | Granturismo | 2018 | -0.411803227701 | 94023.6802545886 | 11 |
| PORSCHE | Panamera | 2018 | -0.373510936023459 | 92142.1739993318 | 14 |
| PORSCHE | Cayenne | 2019 | -0.332623016927105 | 82768.6398230747 | 15 |
| LINCOLN | MKX | 2018 | -0.317051115087406 | 48096.7222329653 | 14 |
| NISSAN | Titan | 2019 | -0.316372969090647 | 49601.6375629378 | 14 |

# Slow Depreciation (...and Appreciation?)

| make <chr> | model <chr> | year <chr> | coefficient <chr> | intercept <chr> | freq <int> |
|---|---|---|---|---|---|
| NISSAN | Altima | 2019 | 0.117760714339714 | 18983.0262054863 | 19 |
| RAM | 1500 | 2019 | 0.100350784445154 | 35048.1436293346 | 15 |
| TOYOTA | Tacoma | 2019 | 0.054238022601273 | 34268.6518782232 | 11 |
| JEEP | Compass | 2018 | 0.0528961499408305 | 21624.8903683508 | 12 |
| JEEP | Gladiator | 2020 | 0.036083900702 | 48683.3474843446 | 20 |
| LAND ROVER | Range Rover Velar | 2019 | 0.011434400609587 | 57611.6311337384 | 12 |
| GENESIS | G70 | 2019 | 0.0065366863953576 | 37302.035963868 | 11 |
| HONDA | Civic | 2019 | -0.00232940467528552 | 23533.2585222751 | 19 |
| GMC | Canyon | 2019 | -0.00358263402109176 | 38349.8520290326 | 12 |
| BMW | X4 | 2019 | -0.00876791825644928 | 48285.6807160439 | 17 |

# Where to next

- So much more to learn from this dataset
- How much do the repair cost?
- Average price of parts and difficulty of repair
  - https://www.rockauto.com/repairindex/
- Expand to new zip codes
  - Do hills put more wear on the powertrain
  - Does rain and snow cause more chassis/frame/undercarriage rust
- All open source and available on GitHub