

Project Milestone 1:

The data set for the term project is about used vehicles listing posted on truecar.com as CSV files, with NHTSA providing detail information of vehicle using an API by providing VIN numbers. Wikipedia site providing emissions policies of each state a vehicle resides and manufacturing locations of USA vehicles in tabular form.

CSV:-

The two csv files contain about 1.3mil listing of used cars as of snapshot date 9/24/2017 in USA.

The headers on the csv files are

Price: Price of the listed vehicle

Year: Model year of the vehicle

Millage: Number of miles vehicle has as of listing date

City: City in which vehicle is located

State: State in which the vehicle is located

Vin: Vehicle identification number listed

Make: Manufacturer of the vehicle listed

Model: Model of the vehicle listed

ID: Unique number for each vehicle(generated)

VIN number is going to be the link to API data listed below. Along with Make for website data.

Cleaning/Formatting Flat File Source:

Model column has scope to separate multiple models in to separate columns and name them appropriately, along with formatting the data elements.

There seems to be vehicles that are repeating and like to perform duplicate search and remove. Along with removing incomplete information of vehicles or that looks abnormally off the normal range (outlier detection and removal) and fuzzy matching.

Make, Model & VIN number is going to be the **join keys** with API and Website data.

The source for the data set is from Kaggle, published by Evan Payne

<https://www.kaggle.com/jpayne/852k-used-car-listings>

API:-

National highway transportation safety association (NHTSA) Product Information Catalog Vehicle Listing Application Programming Interface (API) provides information on Vehicles and their specifications. A call to this API end point with VIN returns detail vehicle information provided by manufacturers to DMV's and to NHTSA.

<https://vpic.nhtsa.dot.gov/api/Home>

For the term project, I am choosing this End points of an API that accepts VIN number as input parameters and decodes the VIN number and returns either a Json or CSV file.

Connecting to an API/Pulling in the data and cleaning/formatting:

The returned json file is in Dictionary form and requires lot of cleaning in terms of removing missing values for attributes in a consistent form. Along with renaming the replacing headers, that would be in more user friendly and is consistent and no missing values. And if missing, would populate default values and does not have nulls.

Below is the API call

```
import requests,json;
endpointurl = 'https://vpic.nhtsa.dot.gov/api/vehicles/DecodeVINValuesBatch/';
post_parameter_fields = {'format': 'json', 'data': 'KL4CJASB6HB002683'};
r = requests.post(endpointurl, data= post_parameter_fields);
print(r.text)
```

Below are detail list of vehicle attributes returned from NHTSA

"ABS": "Standard",	"ActiveSafetySysNote": "",
"AdaptiveCruiseControl": "",	"AdaptiveDrivingBeam": "",
"AdaptiveHeadlights": "",	"AdditionalErrorText": "",
"AirBagLocCurtain": "All Rows",	"AirBagLocFront": "1st Row (Driver & Passenger)",
"AirBagLocKnee": "1st Row (Driver & Passenger)",	"AirBagLocSeatCushion": "",
"AirBagLocSide": "1st & 2nd Rows",	"AutoReverseSystem": "Standard",
"AutomaticPedestrianAlertingSound": "",	"AxleConfiguration": "",
"Axles": "2",	"BasePrice": "24365",
"BatteryA": "",	"BatteryA_to": "",
"BatteryCells": "",	"BatteryInfo": "",

















"BatteryKWh": "",	"BatteryKWh_to": "",
"BatteryModules": "",	"BatteryPacks": "",
"BatteryType": "",	"BatteryV": "",
"BatteryV_to": "",	"BedLengthIN": "",
"BedType": "",	"BlindSpotMon": "",
"BodyCabType": "",	"BodyClass": "Sport Utility Vehicle (SUV)/Multi-Purpose Vehicle (MPV)",
"BrakeSystemDesc": "",	"BrakeSystemType": "Hydraulic",
"BusFloorConfigType": "Not Applicable",	"BusLength": "",
"BusType": "Not Applicable",	"CAN_AACN": "Standard",
"CIB": "",	"CashForClunkers": "",
"ChargerLevel": "",	"ChargerPowerKW": "",
"CoolingType": "",	"CurbWeightLB": "",
"CustomMotorcycleType": "Not Applicable",	"DaytimeRunningLight": "Standard",
"DestinationMarket": "",	"DisplacementCC": "1400.0",
"DisplacementCI": "85.43324173262",	"DisplacementL": "1.4",
"Doors": "4",	"DriveType": "FWD/Front Wheel Drive",
"DriverAssist": "",	"DynamicBrakeSupport": "Standard",
"EDR": "",	"ESC": "Standard",
"EVDriveUnit": "",	"ElectrificationLevel": "",
"EngineConfiguration": "",	"EngineCycles": "",
"EngineCylinders": "4",	"EngineHP": "",
"EngineHP_to": "",	"EngineKW": "",
"EngineManufacturer": "GME",	"EngineModel": "LUV MFI, VVT: Variable Valve Timing, ALUM, E85 MAX ",
"EntertainmentSystem": "",	"ErrorCode": "0",
"ErrorText": "0 - VIN decoded clean. Check Digit (9th position) is correct",	"ForwardCollisionWarning": "",
"FuelInjectionType": "Multipoint Fuel Injection (MPFI)",	"FuelTypePrimary": "Gasoline",
"FuelTypeSecondary": "",	"GCWR": "",
"GCWR_to": "",	"GVWR": "Class 1C: 4,001 - 5,000 lb (1,814 - 2,268 kg)",
"GVWR_to": "",	"KeylessIgnition": "Standard",
"LaneDepartureWarning": "",	"LaneKeepSystem": "",
"LowerBeamHeadlampLightSource": "",	"Make": "BUICK",
"Manufacturer": "GENERAL MOTORS LLC",	"ManufacturerId": "984",
"Model": "Encore",	"ModelYear": "2017",
"MotorcycleChassisType": "Not Applicable",	"MotorcycleSuspensionType": "Not Applicable",
"PedestrianAutomaticEmergencyBraking": "",	"PlantCity": "BUPYEONG",

"PlantCompanyName": "GM Korea",	"PlantCountry": "SOUTH KOREA",
"PlantState": "",	"PossibleValues": "",
"Pretensioner": "",	"RearCrossTrafficAlert": "",
"RearVisibilitySystem": "Standard",	"SAEAutomationLevel": "",
"SAEAutomationLevel_to": "",	"SeatBeltsAll": "Manual",
"SeatRows": "2",	"Seats": "5",
"SemiautomaticHeadlampBeamSwitching": "Standard",	"Series": "Preferred",
"Series2": "",	"SteeringLocation": "Left Hand Drive (LHD)",
"SuggestedVIN": "",	"TPMS": "Direct",
"TopSpeedMPH": "120",	"TrackWidth": "",
"TractionControl": "Standard",	"TrailerBodyType": "Not Applicable",
"TrailerLength": "",	"TrailerType": "Not Applicable",
"TransmissionSpeeds": "6",	"TransmissionStyle": "Automatic",
"Trim": "",	"Trim2": "",
"Turbo": "Yes",	"VIN": "KL4CJASB6HB002683",
"ValveTrainDesign": "Dual Overhead Cam (DOHC)",	"VehicleType": "MULTIPURPOSE PASSENGER VEHICLE (MPV)",
"WheelBaseLong": "",	"WheelBaseShort": "100.6",
"WheelBaseType": "",	"WheelSizeFront": "18",
"WheelSizeRear": "18",	"Wheels": "4",
"Windows": ""	

Planning to use two Wikipedia website (1) locations to get State and facility name of vehicle manufacturing locations in USA by providing **make and model as key** of vehicles from the base CSV. And the other (2) is to provide **State as key** to get emission requirements for that state. This state is from the used vehicles listed state.

Manufacturer, make, model, stat and facility.

Name, light duty gasoline, emissions testing, light-duty diesel, emission testing, heavy duty diesel, emission testing, special notes.

Manufacturer	Make	Model	State	Facility
BMW Group ^[1]	BMW	X3	 South Carolina	BMW US Manufacturing Company
		X4	 South Carolina	BMW US Manufacturing Company
		X5	 South Carolina	BMW US Manufacturing Company
		X6	 South Carolina	BMW US Manufacturing Company
		X7	 South Carolina	BMW US Manufacturing Company
Daimler AG ^[2]	Mercedes-Benz	C-Class	 Alabama	Mercedes-Benz U.S. International
		GLE-Class	 Alabama	Mercedes-Benz U.S. International
		GLS-Class	 Alabama	Mercedes-Benz U.S. International
	Dodge	Durango	 Michigan	Jefferson North Assembly Plant
	Fiat Chrysler Automobiles ^[3]	Jeep	Cherokee	 Illinois
Gladiator			 Ohio	Toledo Supplier Park
Grand Cherokee			 Michigan	Jefferson North Assembly Plant
Grand Cherokee			 Michigan	Mack Avenue Engine Complex (2021)
Wrangler		 Ohio	Toledo Supplier Park	
Ram		1500	 Michigan	Warren Truck Assembly Plant
		1500 Classic	 Michigan	Sterling Heights Assembly Plant

Name	Light duty gasoline	State Emissions				Special notes
		Emissions testing	Light-duty diesel	Emissions testing	Heavy-duty diesel	
Alabama ^[4]		No emissions testing required		No emissions testing required		Alabama maintains "volunteer emission testing"
Alaska ^[5]	all model years old (Anchorage)	Biennially		Exempt		
Arizona ^{[6][9]}	Model Years ≥1987 but more than 3 model years old (Phoenix County Maricopa County)	Biennially	1987–2007 (Pima County/ Maricopa County)	Pima County: 30 percent capacity Maricopa County: 20 percent capacity	Model year <1981: 55 percent capacity Model year 1981–49 percent capacity Model year ≥1950–49 percent capacity	Only state which still conduct emission testing dating back to the 1967 model year unlike other states using EPA classification (or how it conduct test for 1968–present) 1968–testing is for jurisdictions who defined EPA standards for vehicle classification since the 90 model year and beyond automobiles must have an exhaust emission system - PCV system, air injection device, controlled spark module) e.g. the absence of a rolling chassis exemption in some states when an automobile reaches an age limit (as in 25 model years old) the vehicle is exempted or pertaining to motor vehicle laws e.g. classic automobiles legislation and decaying vintage license plates

- 1) https://en.wikipedia.org/wiki/List_of_automobiles_manufactured_in_the_United_States
- 2) https://en.wikipedia.org/wiki/United_States_vehicle_emission_standards

