Intro

Project intro:

- **Client:** One of major truck manufactories in the world.
- **Data:** Millions data points collected from software installed in trucks every day.
- **Problem statement from client:**
  How might we identify and prioritize specific data area to help decisions making in the short to medium term.
- **Project expectation:**
  To identify potential core data areas that need to be addressed in order for client to make best choices in moving forward to the vision 2030.
- **Audience:**
  client's data scientist and connection/solution team
- **Period:** March 15 – April 29, 2021
- **Project Organizer:** Mari Haraldsson (PM in Hyper Island)

The way I took to approach this project:

1. **Quick checked all the datasets I had received.**
2. **Self study across all the channels to gain domain knowledge.**
3. **Picked the area I wanted to explore more.**

I chose to focus on Fuel efficiency for this project, and planned to use machine learning methods to predict miles per gallon(MPG)

Please find my study in following slides.

Sophie Hou
DA22 Hyper Island
hongxia.hou@hyperisland.se

Miles Per Gallon
What matters?

Applying machine learning for fuel efficiency in fleet business

Sophie Hou
DA22 Hyper Island

Data and Challenges

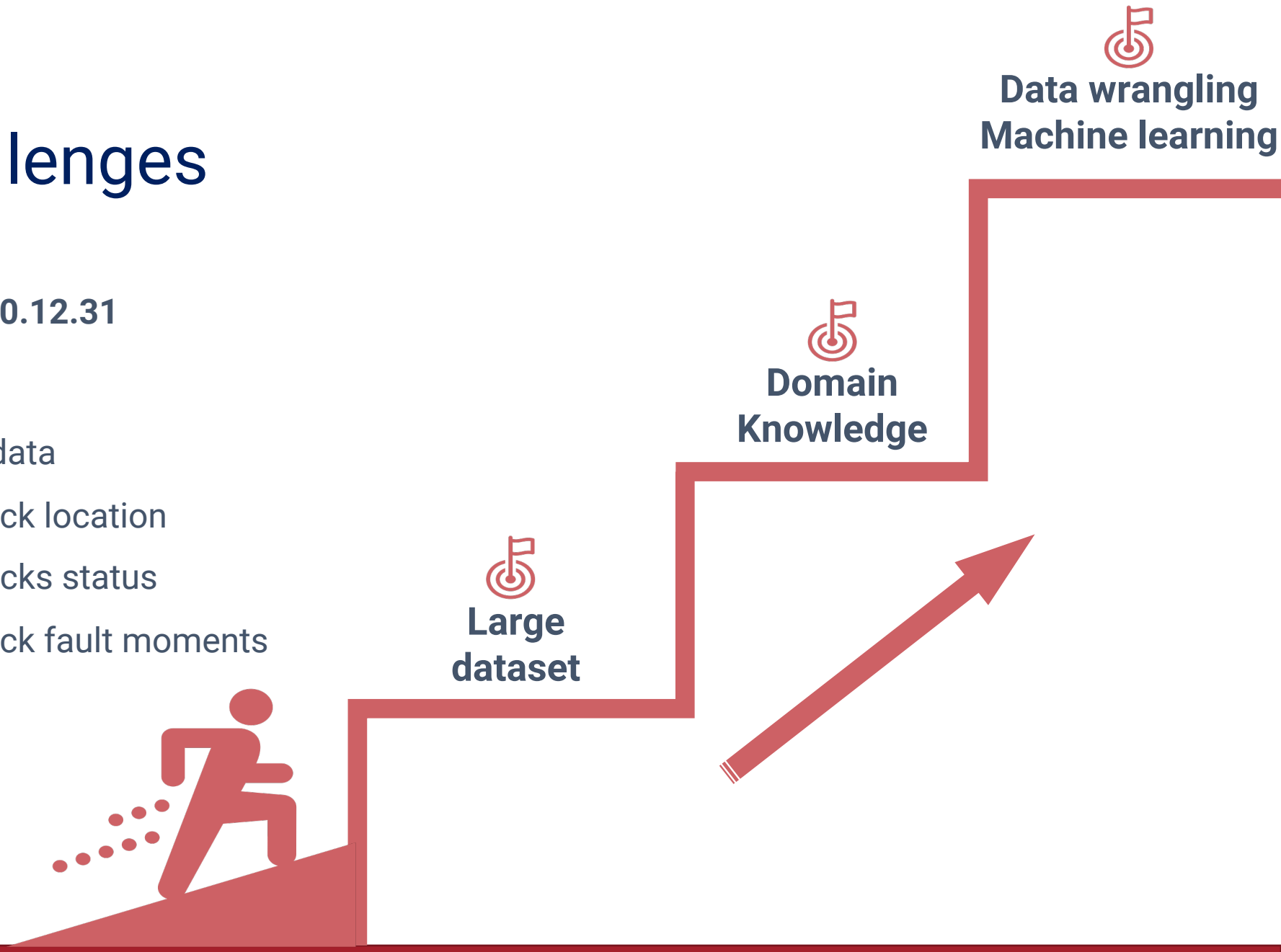
2020-01-01 ~ 2020.12.31

2270 trucks basic data

13 M entries for truck location

27 K entries for trucks status

13 M entries for truck fault moments



The business issues in fleet business

- **Cost control**
  - Fuel
  - Maintenance
- **Safety and Efficiency**
  - Skilled driving.
  - Fuel
  - Uptime

Buy right, repair right, replace right and drive right

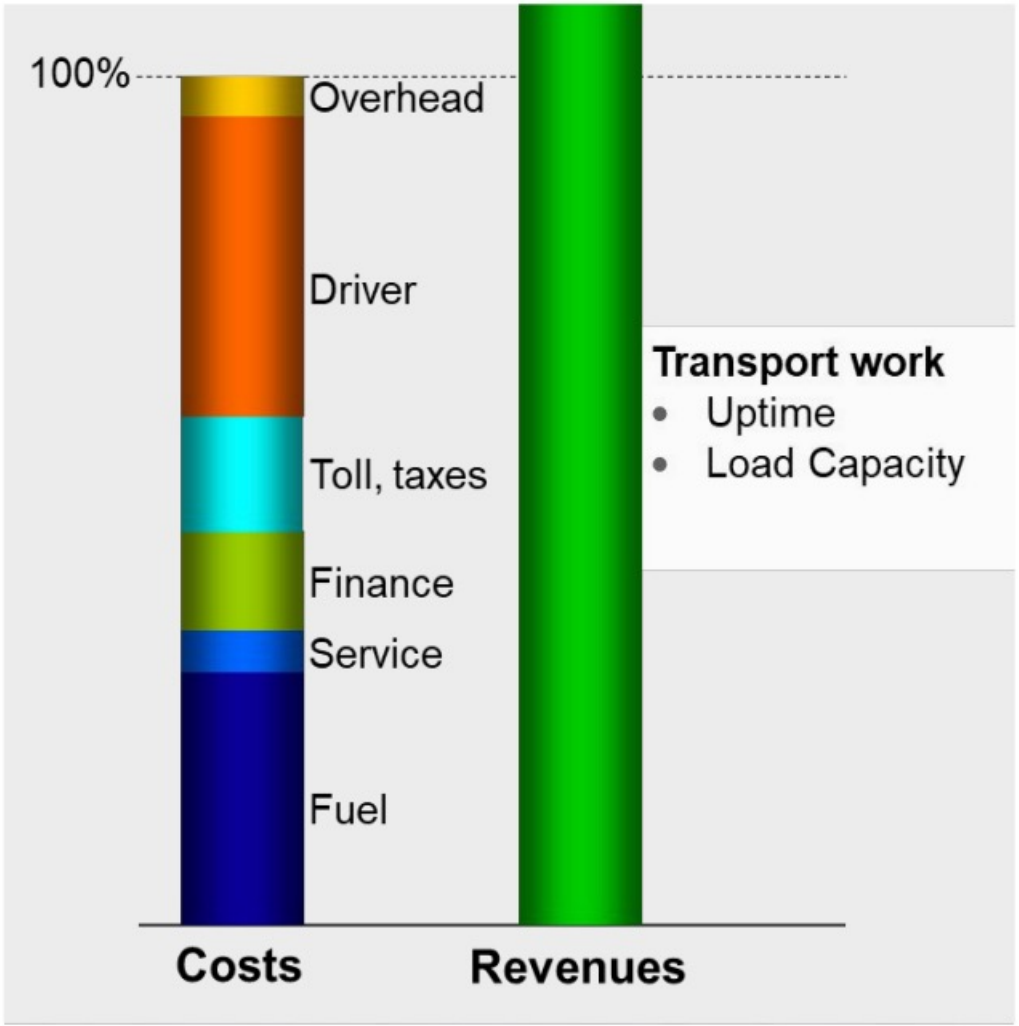


Chart from the client

Questions to answer

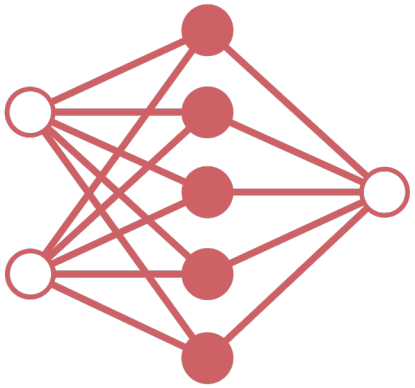
How can data help us to improve fuel efficiency?

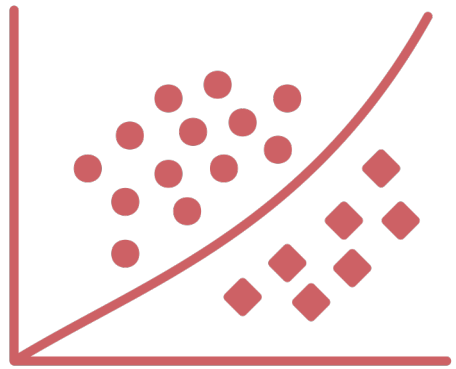
Which machine learning method is better to predict fuel consumption?

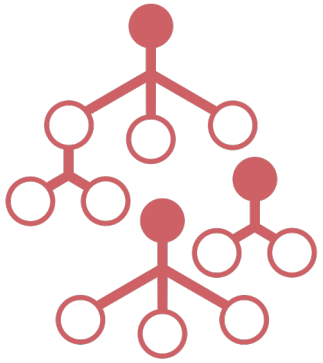
Key Methods to Use

Three type of prediction Machine Learning models

- Support Vector Regression (SVR)
- Random Forest Regression
- Gradient Boosting Regression







Process to take

Choose Data

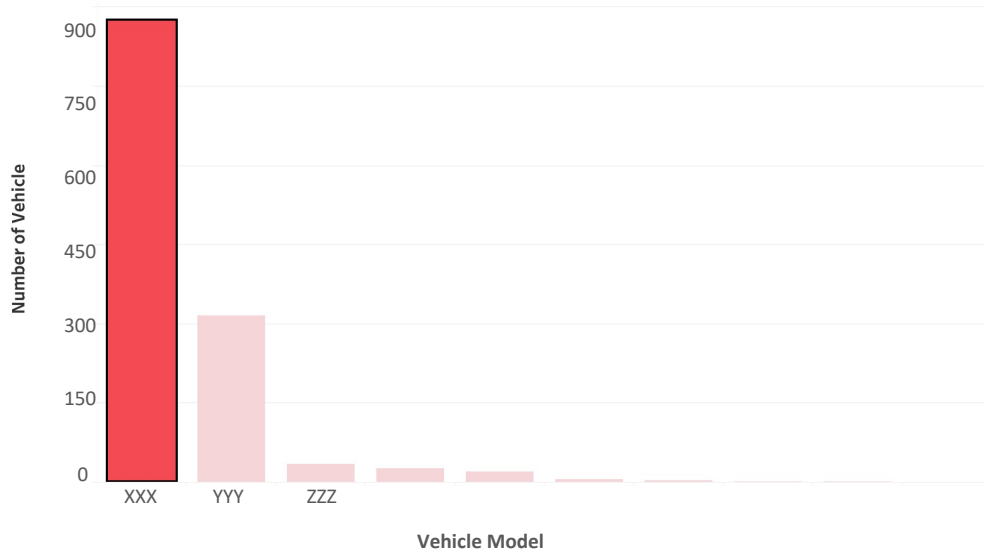
VEHICLE_SPEC

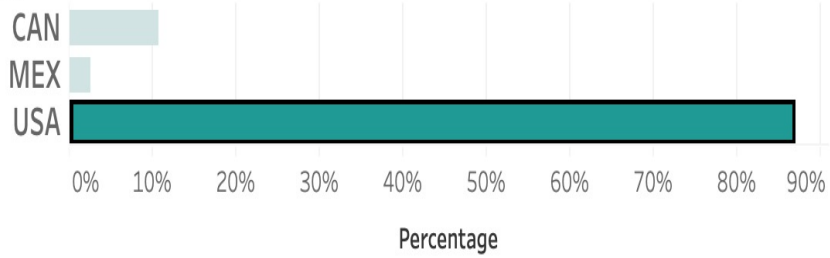
VEHICLE_DETAIL

VEHICLE_LOCATION

Truck Model:
XXX

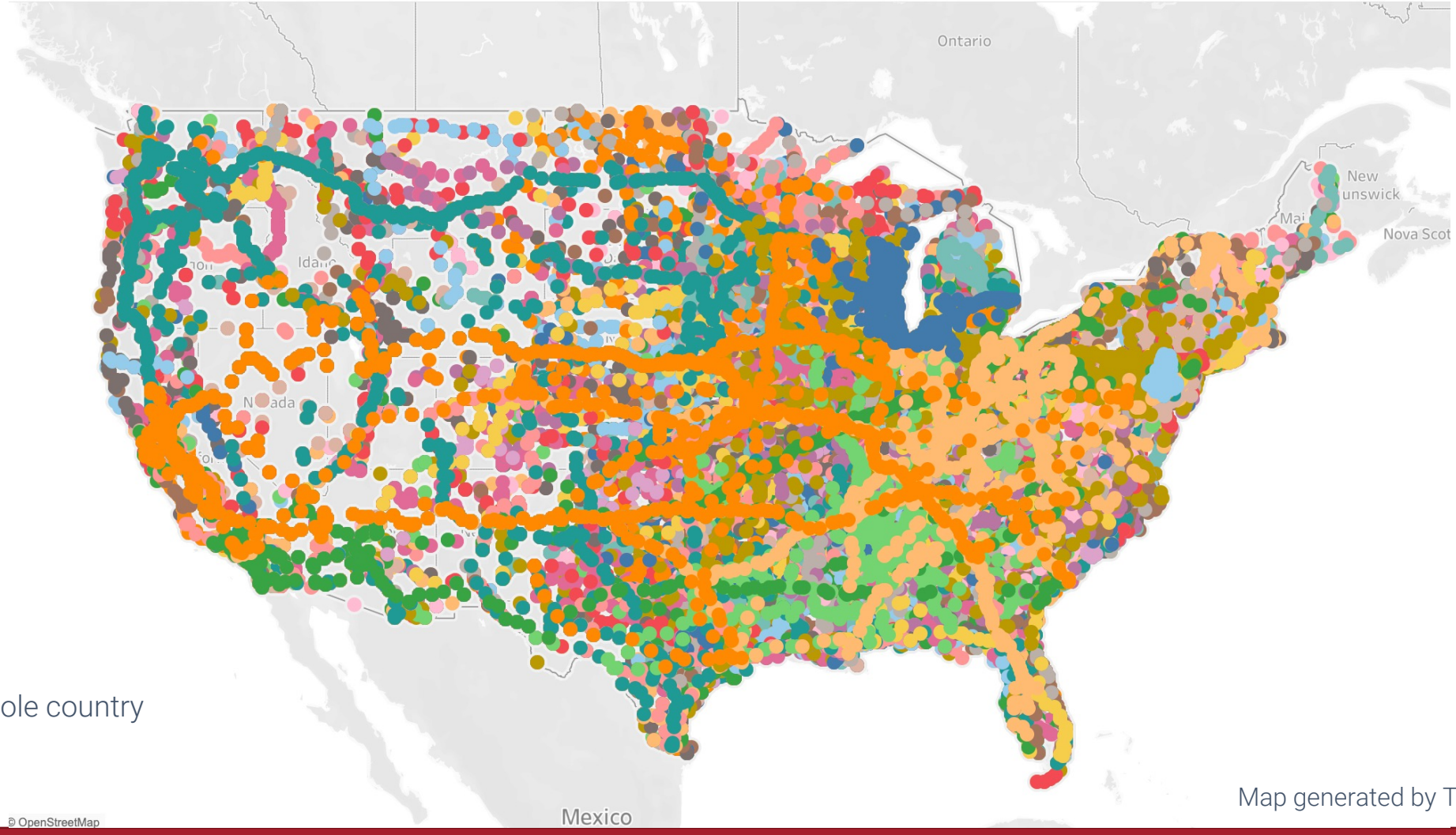
Operation country:
USA





Charts generated by Tableau

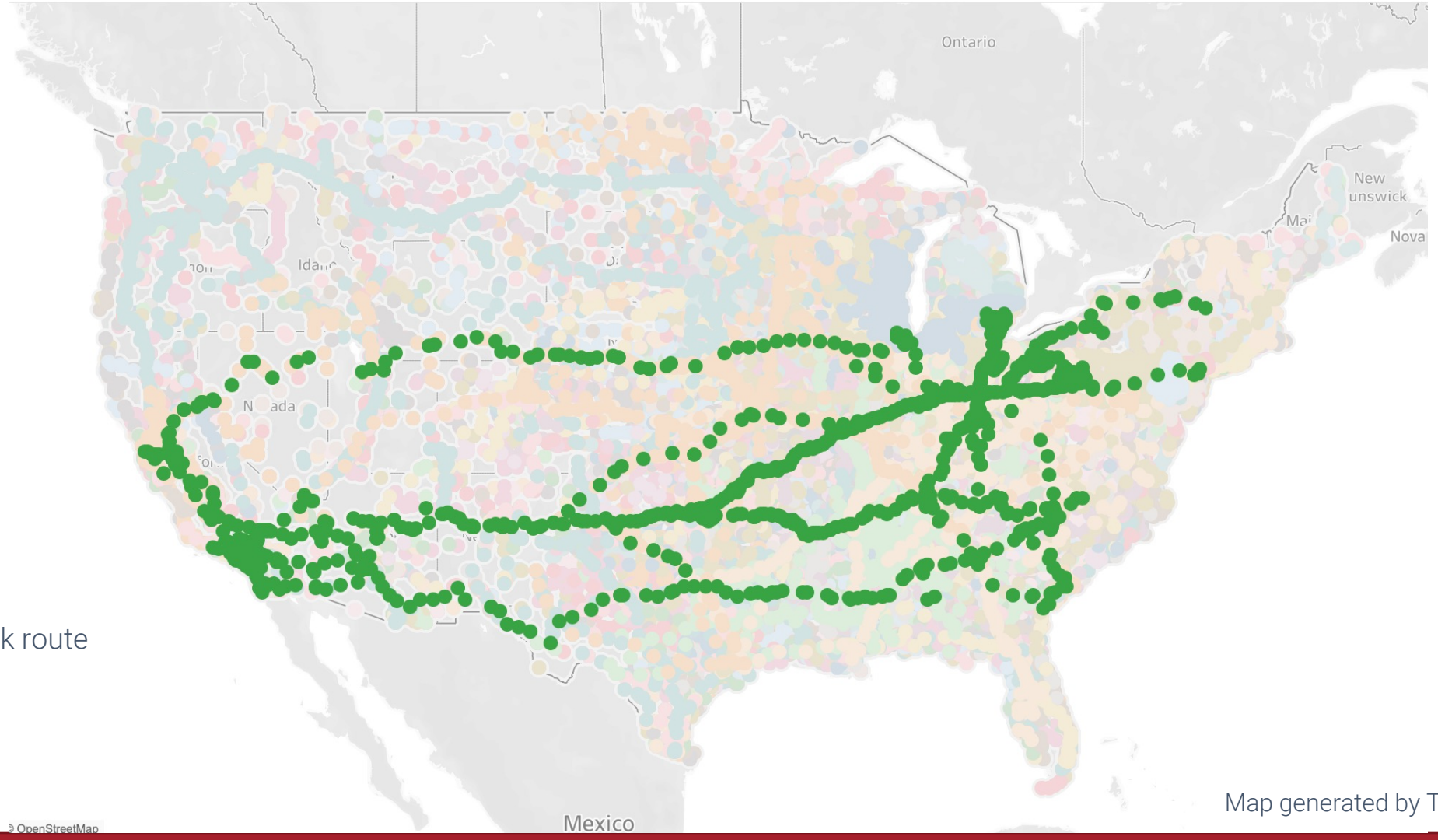
Truck Model -XXX in USA



- Totol 911 trucks
- Long-haul trucking
- Running across whole country

Map generated by Tableau

One truck of model XXX



- Example: One truck route

Map generated by Tableau

Final dataset

Choose Variables,

Cleaning, reshape, calculate to unit per hour,

Join tables, scale, transform, and encoding catergorical variables,

Split to train and test set (70/30)

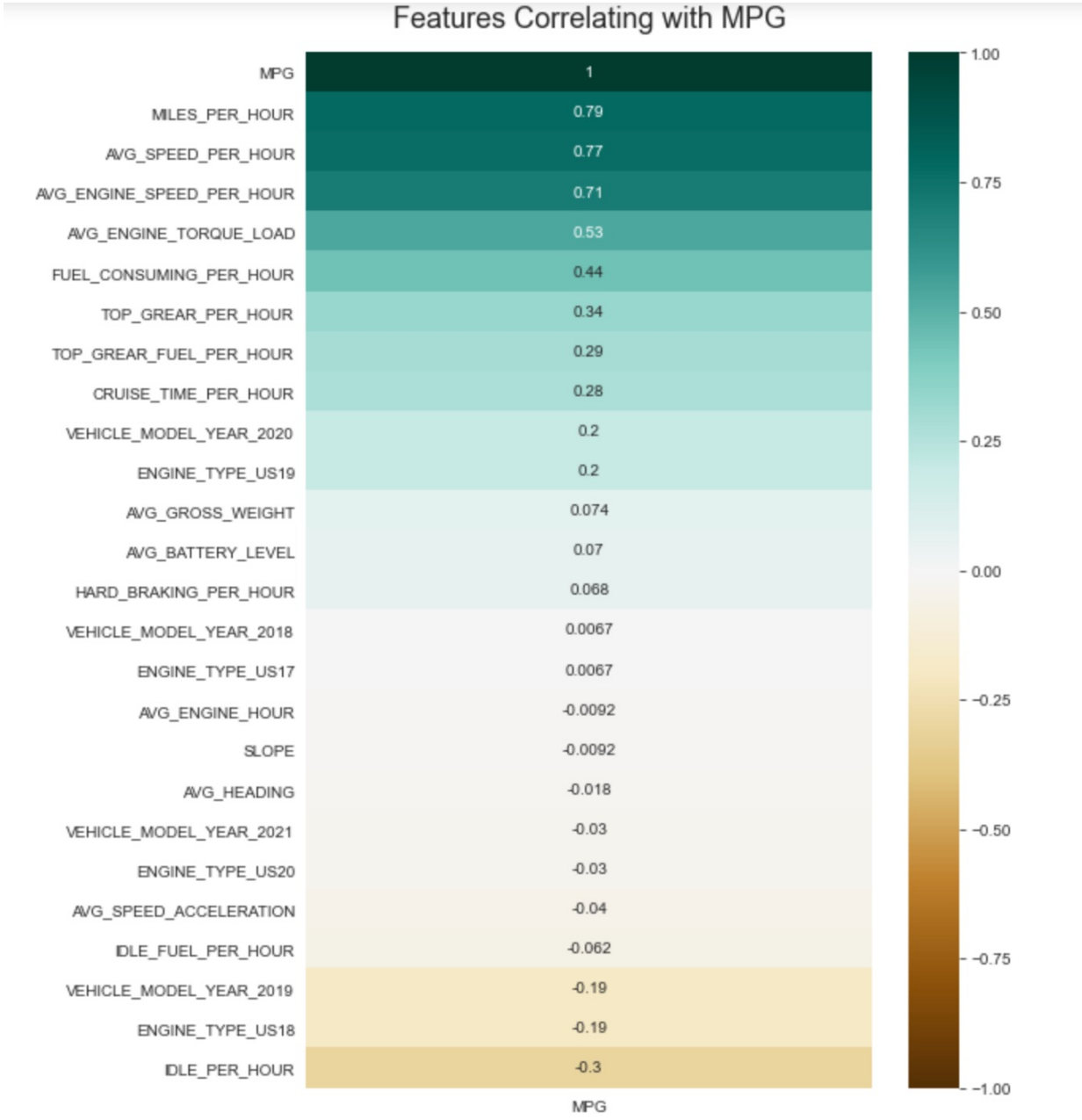
This is what we want to predict

```
RangeIndex: 150951 entries, 0 to 150950
Data columns (total 27 columns):
 #    Column
---   ------
 0    SLOPE
 1    MPG
 2    MILES_PER_HOUR
 3    AVG_SPEED_PER_HOUR
 4    AVG_HEADING
 5    AVG_ENGINE_SPEED_PER_HOUR
 6    AVG_SPEED_ACCELERATION
 7    FUEL_CONSUMING_PER_HOUR
 8    IDLE_PER_HOUR
 9    AVG_ENGINE_HOUR
 10   AVG_BATTERY_LEVEL
 11   AVG_GROSS_WEIGHT
 12   AVG_ENGINE_TORQUE_LOAD
 13   TOP_GREAR_PER_HOUR
 14   TOP_GREAR_FUEL_PER_HOUR
 15   CRUISE_TIME_PER_HOUR
 16   HARD_BRAKING_PER_HOUR
 17   IDLE_FUEL_PER_HOUR
 18   VEHICLE_MODEL_YEAR_2018
 19   VEHICLE_MODEL_YEAR_2019
 20   VEHICLE_MODEL_YEAR_2020
 21   VEHICLE_MODEL_YEAR_2021
 22   ENGINE_TYPE_US17
 23   ENGINE_TYPE_US18
 24   ENGINE_TYPE_US19
 25   ENGINE_TYPE_US20
```

Data info generated from Python

Correlations to MPG

Miles:	0.79
Vehicle speed:	0.77
Engine speed:	0.71
Engine Torque load:	0.53
Top gear:	0.34
Idle time:	-0.30
Cruise time:	0.28



Plot generated from Python

Result of model: Random Forest Regression is more efficient and effective

Prediction methods	Root-mean-square-error	Mean absolute Error	Time
Support Vector Regression	0.768	0.584	long
Random Forest Regression ★	0.181	0.033	short
Gradient Boosting Regression	0.510	0.260	short

Note:

1. Avg MPG(miles per gallon) in test-set: 4.049 mile per gallon.
2. No Gridsearch/cross validation applied yet due to dataset size and project time limitation.
3. Weather influence was not included in this study.

Summary and Suggestion

- Summary from this study
  1. Three regression models to predict miles per gallon.

     Random Forest Regression

     performed better than other two
  2. Engine features and driving behavior strongly associate to fuel efficiency.

- Suggestions for further study:
  1. Extra data collection:
     - Weather Data (temperature, wind, humitity, rain/snow, air pressure)
     - Engine data (Rev, horsepower, size..)
  2. K-means clustering to look for patterns in data for future business decision related to MPG.

Thank you!

Sophie Hou
Hyper Island DA22
hongxia.hou@hyperisland.se

April 2021