

Big Data Analytics: How Big Data is Shaping Our Understanding of Electrified Vehicle Customers

N. Khalid Ahmed and Jimmy Kapadia
 Ford Motor Company

ABSTRACT

Electrified vehicles including Battery Electric Vehicles (BEVs) and Plug-In Hybrid Vehicles (PHEVs) made by Ford Motor Company are fitted with a telematics modem to provide customers with the means to communicate with their vehicles and, at the same time, receive insight on their vehicle usage. These services are provided through the "MyFordMobile" website and phone applications, simultaneously collecting information from the vehicle for different event triggers. In this work, we study this data by using Big Data Methodologies including a Hadoop Database for storing data and HiveQL, Pig Latin and Python scripts to perform analytics. We present electrified vehicle customer behaviors including geographical distribution, trip distances, and daily distances and compare these to the Atlanta Regional Survey data. We discuss customer behaviors pertinent to electrified vehicles including charger types used, charging occurrence, charger plug-in times etc. Throughout this discussion, we highlight the process of extracting information from this data that can be used to further refine future design. Customer data privacy concerns are also addressed, with no analysis using any personally identifiable information.

CITATION: Ahmed, N. and Kapadia, J., "Big Data Analytics: How Big Data is Shaping Our Understanding of Electrified Vehicle Customers," *SAE Int. J. Mater. Manf.* 10(3):2017, doi:10.4271/2017-01-0247.

INTRODUCTION

Big data is a relatively new trend that is influencing the decision making process across various industries. Within the automotive industry, large quantities of data have been historically recorded from vehicle tests to marketing and sales events. Ford Motor Company has been a leader in collecting data and understanding vehicle performance in relation to customer expectations. There has been a recent trend for collecting data that helps understand the interaction of customers with their vehicles. This typically includes but is not limited to various powertrain systems and features. With the advent of affordable cellular technology, this has been taken to the next level where vehicle data is directly uploaded to secure cloud based servers within Ford Motor Company [1]. As a first step, this technology was introduced in the form of 2G modems on certain electrified powertrain vehicles starting in model year 2013. Using this data, in this paper we present findings that focus on average customer behaviors driven by electrified vehicles, rather than the operation of the powertrain in these vehicles.

Electrified powertrain vehicles are a natural choice for collecting vehicle data, as this data is useful to both customers and manufacturers. These vehicles have an electrical energy input and path to the wheels, either in place of or in addition to, the gasoline fuel energy and mechanical path in conventional internal combustion engines [2]. These vehicles include (a) pure electric vehicles known as Battery Electric Vehicles (BEVs) - for e.g. Ford Focus Electric, Nissan Leaf etc.; (b) a combination of electrical and mechanical powertrains known as Fully Hybrid Electric Vehicles (FHEVs) - for

e.g. Ford Fusion Hybrid, Ford C-Max Hybrid, Toyota Prius etc.; and (c) a combination of BEVs and FHEVs known as Plug-In Hybrid Electric Vehicles (PHEVs) - for e.g. Ford Fusion Energi, Ford C-Max Energi, Chevy Volt etc. The specific powertrain architecture of these electrified vehicle types can be different from each other.

Electrified vehicle customers are often anxious about the state of their vehicle, including the battery state of charge, available electric miles, vehicle energy consumption etc., among other pieces of information. By providing customers with this data, customers have more realistic expectations regarding their vehicle's abilities. We also use this data to train customers to drive their vehicles more efficiently. Additionally, customers also benefit from the data collected by using it to plan a trip with charge station information and duration of charging among other parameters.

Big Data

With approximately 33.6k BEV and PHEV customers, who average approximately 2.5 trips a day, the data being collected is voluminous and thus falls into the definition of big data. This enables us to understand behaviors far beyond previous studies [3]. Specifically, we collect a few hundred gigabytes of data per vehicle per year. This includes over three billion data records comprising of over 300 million miles that have been logged. Analyzing this data and performing calculations on the data is beyond the capabilities of a single machine. Thus we use typical big data technologies to address this problem, illustrated in [table 1](#). The data is stored using the

Hadoop Distributed File System (HDFS), while analysis is performed using a number of tools including Hive and Pig tools and Python scripts. We describe this in more detail in the methods section.

MyFordMobile Connectivity (MFM)

The 2G modem on our BEV and PHEV electrified vehicles interfaces with customers through the MyFordMobile (MFM) application [1]. MFM is a smartphone app, website and connectivity feature that enhances the electrified vehicle ownership experience for Ford customers.



Figure 1. Implementation of MyFordMobile connectivity. This figure shows a customer communicating to his electrified vehicle, a Ford Focus Electric, through cellular technology. The customer connects to a Ford cloud-based secure server which in-turn communicates with his vehicle.

MyFordMobile uses a 2G telematics modem integrated into the vehicle allowing the car to communicate with the Ford cloud-computing service through familiar cellular technology, as shown in Figure 1. MyFordMobile allows customers to communicate with the Ford cloud-based secure server to view and monitor their vehicle’s battery state of charge, get updated charge station information and remotely access their car from anywhere at any time.

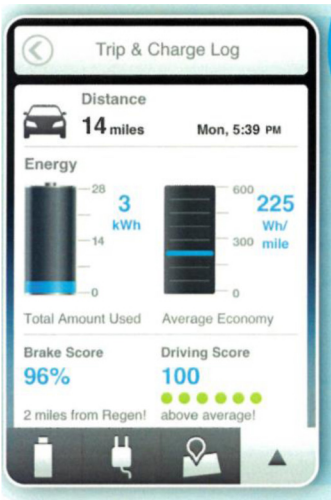


Figure 2. Trip summary statistics. This figure shows a sample of trip summary statistics that are immediately available at the end of each trip. These statistics are uploaded to Ford cloud-based secure servers, allowing for real-world analytics to be performed on electrified vehicles.

This telematics unit on our BEVs and PHEVs gathers metadata after certain events and uploads this metadata about the events to our cloud-based secure server. For example, using this technology, trip

summary statistics of every trip taken by every customer are uploaded to the Ford cloud-based secure servers. As shown in figure 2, these statistics include the total trip distance, electric energy used, gasoline fuel used, trip distance driven with the engine off among other parameters. These statistics are received at the end of each trip, providing data from trips that include different geographical locations, different temperatures and seasons, and different customer driving habits. In addition to trip based data, other event based data include charging events and certain faults.

METHOD

Technology

We use Big Data technologies based on Hadoop Distributed File System (HDFS) [3] to perform various analytical calculations in this work. It is important to realize the difference between presenting data collected and learning customer insights from the data. Throughout this paper, our focus is on the latter as we aim to perform big data analytics with the aim to understand customer behaviors. With this in mind, we pose various questions and then explain our approach to answer these questions in the next section.

Table 1. A brief description of various big data technologies used in this project.

Hadoop	Data storage
HiveQL	Exploratory analysis to lay the analytics foundation to be performed on all the data
Pig Latin	Parallel analysis of data on the Hadoop File System (HDFS)
Python	Perform analytical calculations on a local machine that can be streamed upward and parallelized
QlikView	Visualize data by building useful dashboards that can be queried easily

Big data analytics involves the use of new technological tools and programming paradigms to find answers hidden in big data. In table 1, we present a brief description of the various big data tools used in this project. Data obtained from PHEVs and BEVs is stored in a Hadoop Distributed File System (HDFS) [4]. As a first step, exploratory analysis is performed on this data using Hive Query Language (HiveQL) [5]. The aim of this step is to understand the data available to answer a given question and determine the analytics to be performed. As a next step, we use Pig Latin scripts that are set up to perform the analytics in a parallel fashion on the HDFS system. Pig Latin is a programming language that uses Pig, a high-level platform for creating programs that use Map-Reduce to run on the HDFS system [4]. In addition to using Pig, we also perform a large number of analytical calculations in Python [6] on a local machine using pandas [7], numpy, scipy [8], and scikit-learn [9] packages among others. These open-source packages are readily available online and

assist in a number of ways, including pandas for data analysis in python through the implementation of dataframes, numpy for performing numerical calculations on python objects, scipy for more complicated numerical calculations including fitting various functional forms to data and scikit-learn to perform machine learning tasks such as finding natural clusters in the data. The goal of this step is often to develop new calculations and functions that can be streamed upwards to the HDFS and then parallelized within the Pig framework. Finally, we use Qlikview [10] as a tool to create dynamic reports by building useful dashboards on a server where possible. These dashboards can then be queried easily to modify the way any piece of information learnt is conveyed to the audience.

Event Based Data Collection

In this project, we receive snapshots of data at the end of trigger events on our PHEV and BEV electrified vehicles. This is initially setup through a rigorous process to ensure the safety and security of our customer data. This process includes an authorization on two different instances in time that are at least 24 hours apart, between the vehicle and a mobile device. Once the required authorization has been received, we receive event based data including trip metadata, charger plug event information, a charge full alert, a pre-conditioning alert, a vehicle alarm alert, a charging fault alert etc. These pieces of data are valuable to our customers - for example, a charge full alert informs a customer that his/her vehicle is fully charged and ready for a trip or maybe moved out of the charging station. Pre-conditioning, which is the use of electricity from the grid to prepare the cabin for the next trip, can be setup from a customer's mobile device. The vehicle then communicates to the customer that pre-conditioning was successfully completed. This information is also transmitted to the Ford cloud-based secure server and can be used to understand various trends that might exist in the pre-conditioning usage among our customers.

We wish to emphasize that customer privacy is of utmost importance in our work. In this project, we utilize only non-personally identifiable information for the purpose of performing analytics on the user's experience and vehicle performance. Additionally, this analysis is performed on an aggregate level and does not identify any customer or his information personally. Rather, non-personally identifiable information collected through the MyFordMobile service is used to compile aggregate and statistical data to help vehicle design and identify popular features and thus make our customer experience friendly. We eventually hope to use this aggregate information about our customers to develop new and improved products and services.

RESULTS

In this section we present results to various questions posed on customer behavior and other information from MyFord Mobile data. It should be noted that none of the data here is from surveys and the like, thus removing any subjectivity in the numbers shown. Rather all the data is collected electronically with a very high accuracy.

Descriptive Analytics

As a first step, we present a few descriptive analytics about the vehicle in the MyFord Mobile dataset. In figure 3, we show the number of unique vehicles available in this dataset as of August 2016. There are approximately 16.6k Fusion Energi PHEVs, 12.2k C-Max Energi PHEVs and 4.7k Focus BEVs.

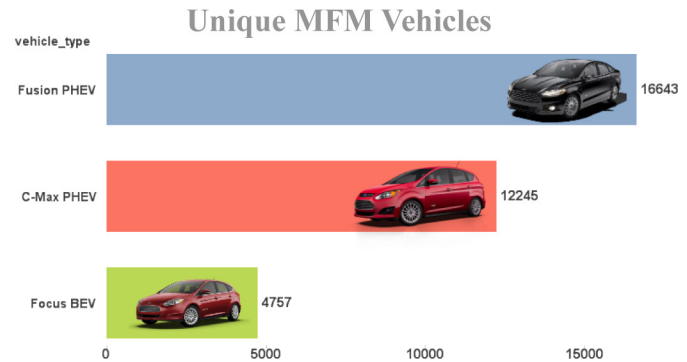


Figure 3. Unique vehicles in MyFord Mobile. We actively receive event based data from approximately 33.6k unique PHEVs and BEVs.

As explained in the previous section, we receive event based data from over 33.6k unique PHEVs. In figure 4, we only consider the trip summary data and plot the number of unique trips over time. We observe a growth in the number of unique trips each year, as the number of unique vehicles has grown over time. As of the writing of this paper, we had data available to us until August 2016. To guide the eye, we have drawn a line connecting the months of August from each year starting in 2013.

We would also like to point out that in this study, we only consider PHEVs and BEVs of 2013 - 2016 model year vehicles which are fitted with a 2G telematics modem. Later model year vehicles have a 3G telematics modem and are not included in this study.

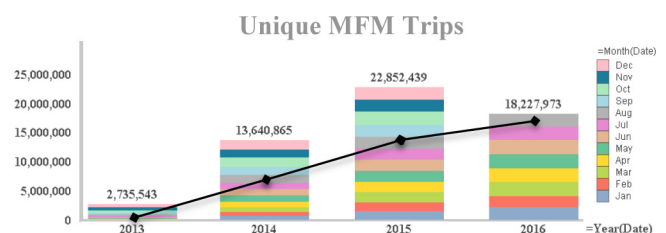


Figure 4. Steady growth in the number of unique trips over time, due to increase in customers. The black line shows the growth up to the month of August each year, as a guide to the eye.

We also present the geographical distribution of these vehicles in figure 5. As can be observed from the figure, approximately 24% of the vehicles are in the states of California and Nevada. A large number of the vehicles are also in the North Eastern states, while Michigan accounts for approximately 3.25% of the vehicles. Note that the density bubbles in figure 5 are centered on the State labels. Understanding this geographical distribution of the vehicles helps determine the extreme weather conditions that the vehicles need to be designed against [11].