Qualetics
DATA MACHINES

Usecase

Time Series Forecasting to Analyze LPG Usage

Digital Intelligence in Energy Sector

Introduction:

Liquefied Petroleum Gas (LPG) is the preferred choice of fuel across the world. Increasing usage of LPG can be attributed to its affordability, efficiency, and environment-friendliness. India's demand for LPG is continuously rising with the government measures to provide cleaner cooking gas even in rural areas. Indians consumed a record 24.9 million tonnes of LPG in 2018-19 and analysts are expecting the consumption to keep rising.

However, unlike in many countries where LPG is supplied through pipelines, India delivers LPG via individual cylinders booked by consumers. The supply of LPG through pipelines allows for constant metering and measurement of usage, whereas an LPG supply through a cylinder doesn't provide monitoring of activities like usage, safety, and service quality.

Sensors, Internet of Things (IoT), and AI-powered software enable companies in digesting large amounts of data and provide real-time responses. The current case study leverages AI and Machine Learning to predict LPG usage from different perspectives by using a concept mechanism like a trolley. These trolleys capture the weights of the cylinders and transmit continuous updates on weight of the cylinder, gas leakage occurrences and ambient temperature to the dealers and manufacturers. Having the ability to capture the above-mentioned data points continuously will allow the possibility of real-time streaming analytics of the LPG gas usage as well as advanced analytics on data captured over long periods of time.

> "Offshore platforms are, on average, running at only 77 percent of maximum production potential. Rigorous use of analytics can improve substantially on that."
>
> -Mckinsey & Company

Problem Statement/Challenges:

With increasing consumption of LPG, there arises a need to meet the demand-supply requirements. Without proper data, it will be difficult to achieve it. Like other energy sources, there are challenges that disrupt the smooth supply of LPG. Proper and structured data gathered from sensors and other sources will empower the manufacturers, dealers, and consumers with actionable insights on consumer usage patterns.

The intent of the current study is to analyze the LPG usage consumption and forecast the values to predict the LPG usage - by giving inputs area-wise, dealer-wise, and season-wise on a weekly, monthly, and yearly basis.

About the Dataset:

We obtained a dataset that consists of two different files.

Dealers Dataset - This dataset has recorded entries about the information from the dealer which covers the number of cylinders supplied each day from 2013 to 2017.

Consumer Dataset -This dataset has recorded entries relating to the weight of LPG measured in grams from the year 2018 to 2019. This data is only recorded for 3 consumers from dealer1. There were 2 sub files in this dataset that have data on weekly usages that were recorded for each minute and monthly usages that were recorded for each day.

Solution provided:

Looking at the consumer dataset, it was observed that the recorded entries were only for one week's timestamp. It becomes difficult to use such a dataset for future forecasting as the data was recorded per second. Hence, we restricted the usage of this data to our analysis and went ahead for future predictions using the monthly data as this data was recorded for each day.

Time series data can reveal a variety of patterns. It is beneficial to split the time series into several components, each component representing an underlying pattern category. To get a deeper understanding of the time series as well as to improve our forecast accuracy, we decomposed the data. Time series decomposition often involves deconstructing a series as a combination of the average value of the series (level), the increasing or decreasing values in the series (trend), the short-term cycle in the series (seasonality), and the sequence of random and unpredictable values (noise) components.

We decomposed our data into trend, seasonality, and residual charts. The data consists of white noise, which implies that there is much randomness present in the data which will affect our analysis and prediction. Therefore, we forego the seasonality and trend components.

As there was too much randomness in the data, we opted for neural networks.

Technologies Used:

Data Preprocessing

Time Series Analysis

Neural Networks

Our Process:

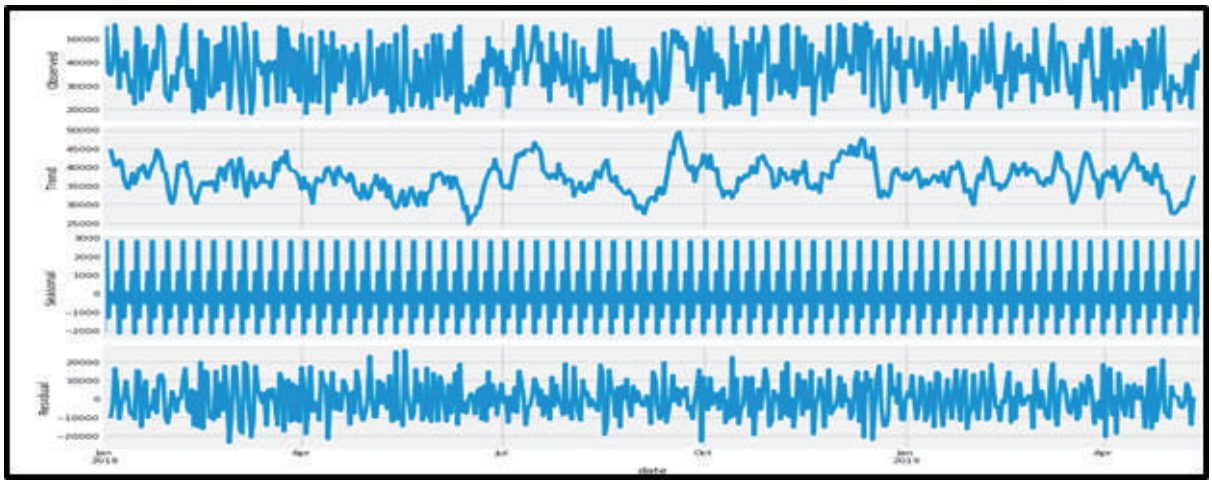
Neural Networks are extremely useful in identifying patterns that are too complex to be manually extracted and taught to the machine. These neural networks are computing systems that consist of simple and highly interconnected elements, called "neurons". These neurons help process information using dynamic state responses to external inputs.

In Neural Networks, we used Long Short-Term Memory (LSTM) for our study. LSTMs are considered as the most effective solutions for prediction problems as they have an edge over the conventional neural networks and artificial Recurrent Neural Network (RNN). LSTMs are enabled with a unique feature of selectively remembering the patterns for long durations of time.

Using LSTM, we tried to create a time series model that was able to learn the pattern and accordingly predict the LPG usage consumption values.

The alternate model used was SARIMA, which is one of the most used techniques in time series forecasting. SARIMA is the model which is capable of capturing trend and seasonality.
We used LSTM to evaluate the working of SARIMA. Further, we took a cumulative of all the days for each month and monthly usage data and we predicted the future demand for 5 months.

Decomposition chart for daily-usage data



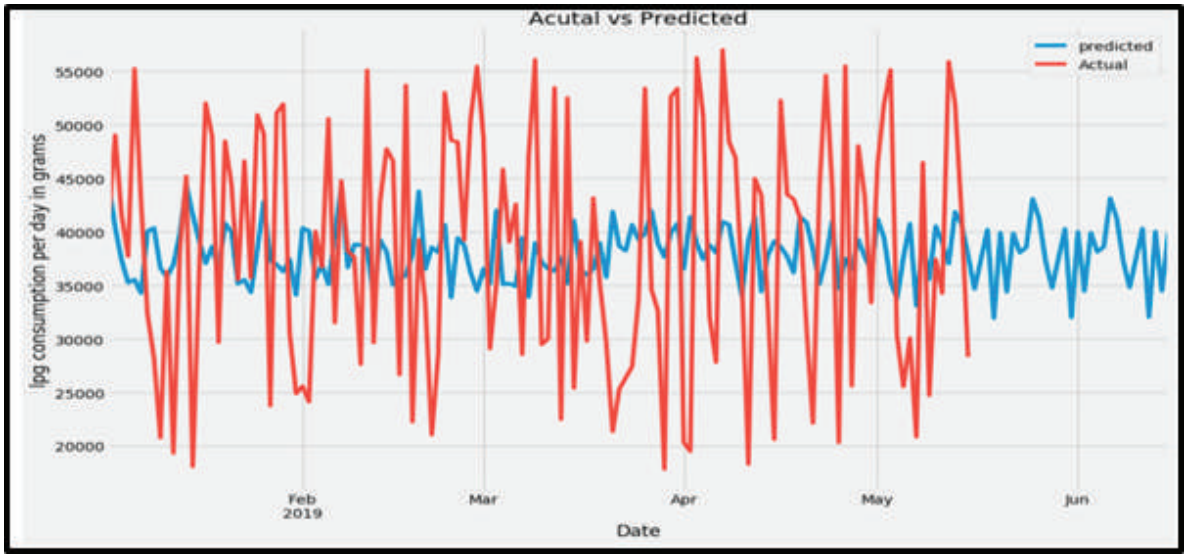
The plot clearly shows that data has white noise present and implies high residuals and randomness. i.e no seasonality and trend. (This is due to data fabrication).

White noise - sound waves extending over a wide frequency range

Seasonality - periodic fluctuations repeated over time

Trend - direction of change

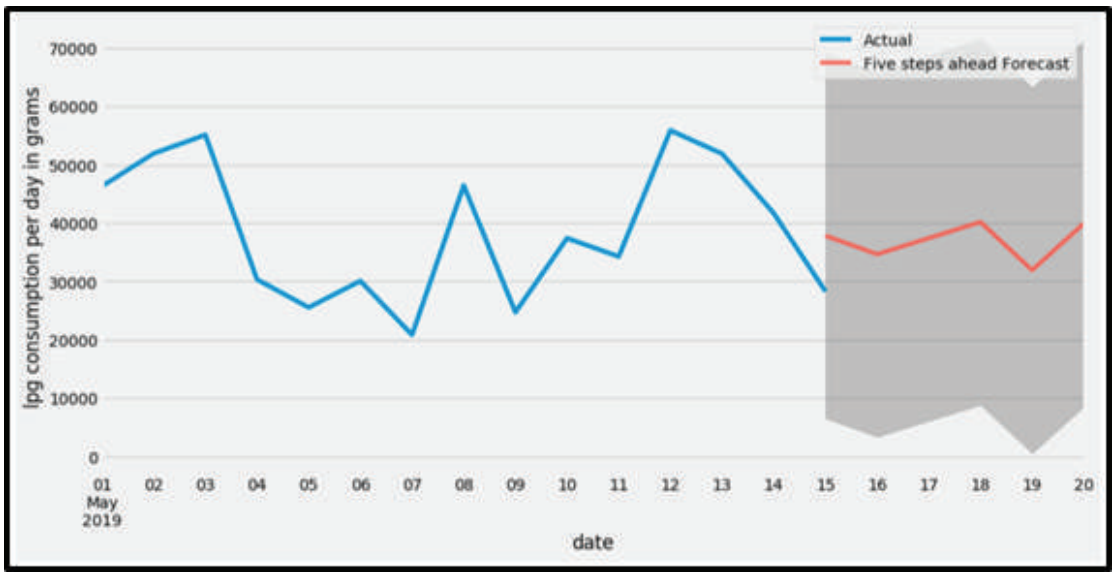
Daily-usage analysis using the SARIMA model



This plot denotes the model's predicted values over the actual values for each day.

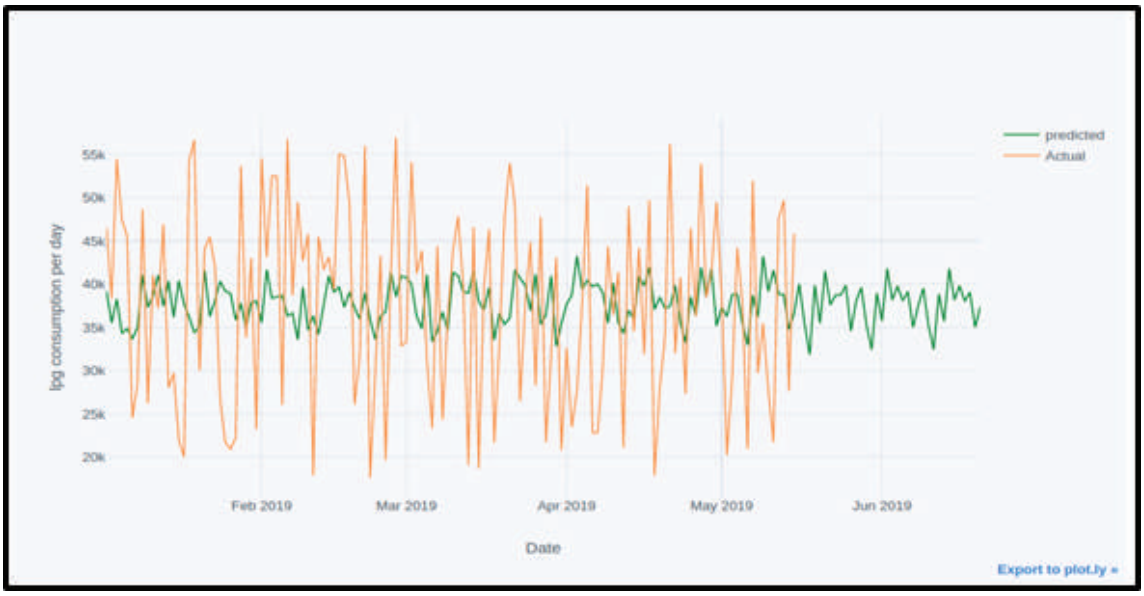
(Note: Red line represents the actual consumption of LPG by the dealer and the blue line represents the forecasted values of consumption, Y-axis is the LPG consumption in Grams)

Future prediction for daily-usage using time series analysis model SARIMA



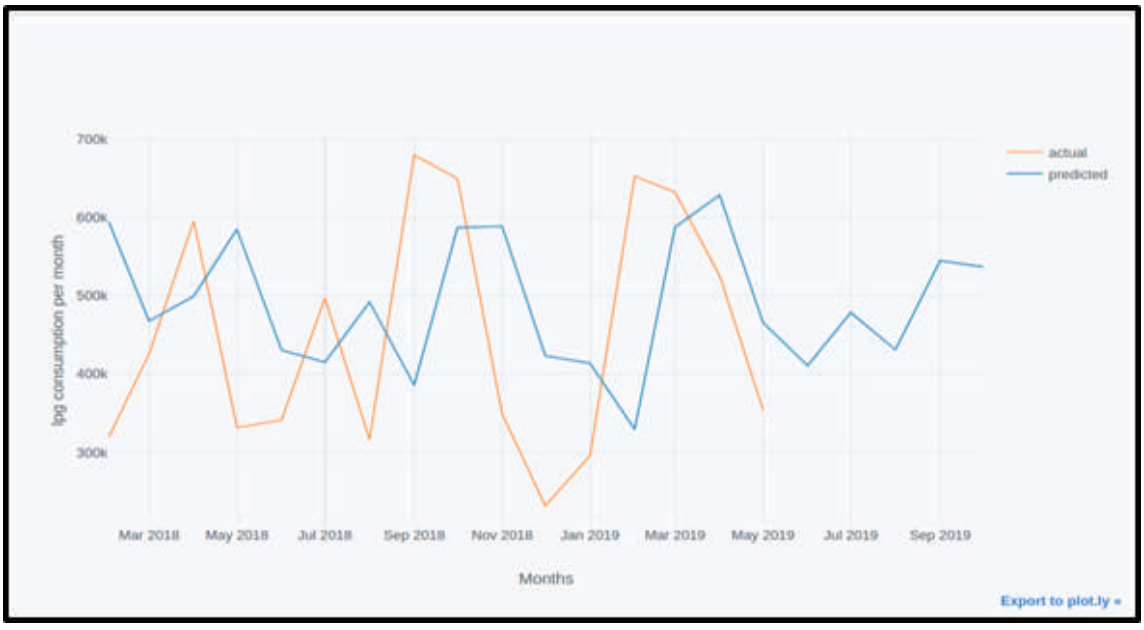
We predicted the LPG usage consumption for 5 days i.e., from 15th May 2019 to 20th May 2019. The blue line is the actual consumption and the red line is the prediction for 5 days mentioned.

Daily-usage analysis using the Neural Network model



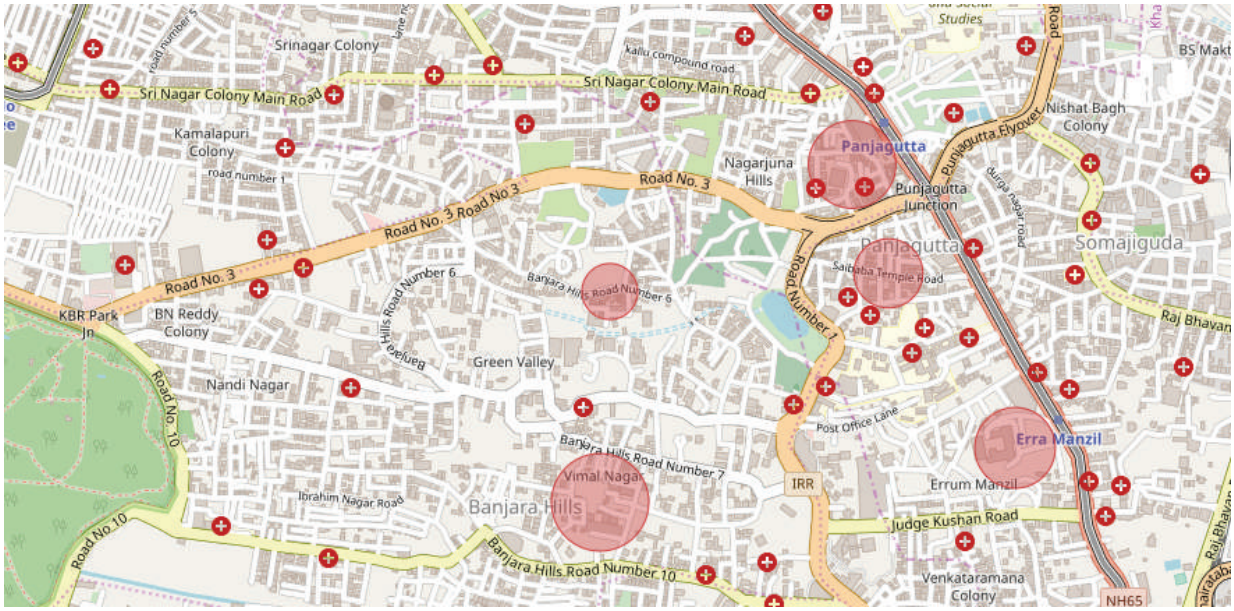
This plot shows us the model's predicted value over the actual values for each day along with the forecasted values. The actual data is represented with orange and the predicted data with green.

Monthly-usage data using the Neural Network model



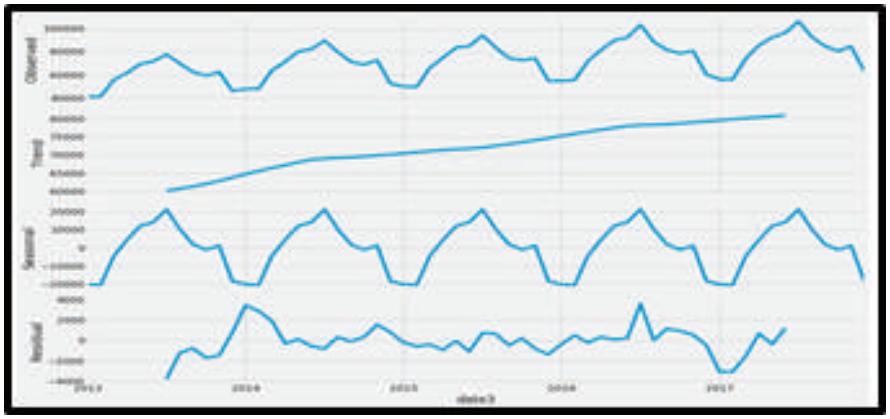
This plot shows the model's predicted value over the actual values for each month., where the orange line depicts the actual value and the blue line depicts the predicted value along with the forecasted values. Y-axis is the LPG consumption measured in Grams.

Time Series Analysis on Dealers Data Set

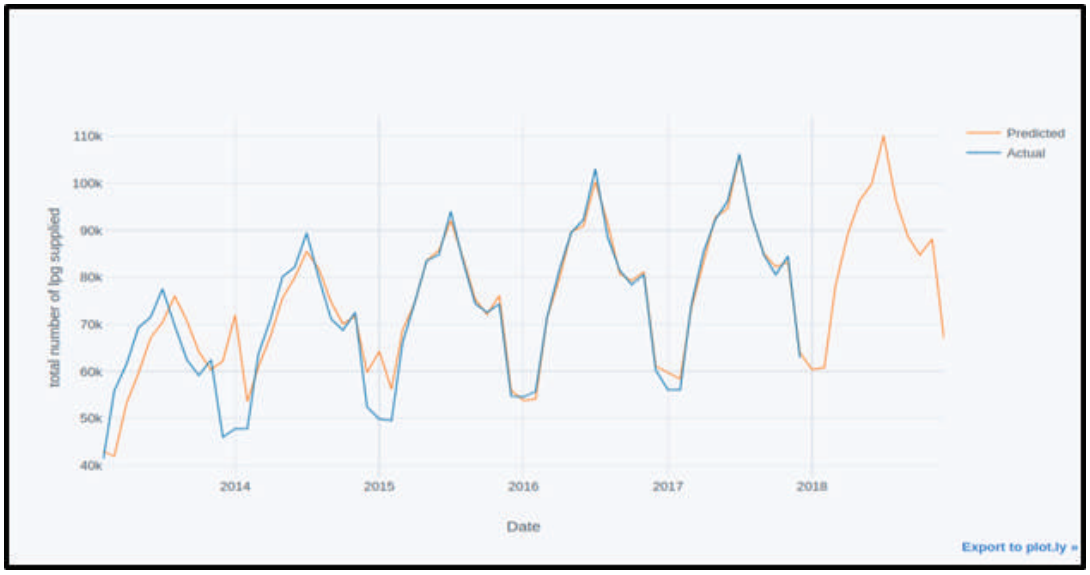


This plot shows the areas to which the dealer supplies cylinders, represented by the blue icons. The analysis gives insights on the reachability of a particular dealer and to which other localities the dealer can supply the cylinders and increase their reach.

Decomposition chart on dealers dataset



To understand the number of cylinders sold, an analysis was performed on the dealers' dataset. We performed the decomposition technique on it. The above plot infers that there is a linear trend present in the data set with high residuals.



This plot depicts the behavior of the model used denoting the actual vs predicted values. Blue line is the actual value in our data, orange line is the predicted values by the model over the actual values and also includes the forecasted values for 2018.

Model Metrics:

Time series is an important area of Machine Learning as it can be used to understand and analyze the past as well as predict the future.

We chose to go with 2 models to determine the accuracy of our time series forecasting performed in our analysis. The accuracy levels of achieved on each data set are shown in the tables below:

For consumer data set:

	SARIMA	LSTM
Accuracy (for daily usage)	59%	76%
Accuracy (for monthly usage)	Not used	81%

The table clearly indicates that the Neural Network performed better as compared to SARIMA as the data consists of noise. Our approach to taking a cumulative of the month helped to increase the accuracy as well.

For dealer data set:

	SARIMA	LSTM
Accuracy	74%	89%

Likewise, the same approach was used for the dealers' dataset. It was observed in the decomposition chart that the data had a high trend. LSTM performed relatively better because there are outrageous residuals present in the dataset.

Qualetics as a Solution:

Qualetics provides the dealers and manufacturers with a portal that will provide them with the dashboard consisting of insights related to LPG usage by the consumer. Consumers' usage data can be integrated with the dashboard with the help of sensors and user-app. The goal of this case study is to analyze and understand the supply and demand requirements for residential usage of LPG at a dealer or manufacturer level as well as consumer level. This analysis helps the various parties involved in planning their future course of action.

The benefits derived from our study can be summarised into the table below:

Dealers

- Track present sales
- Predict future demand
- Track the service team performance and draw actions for improvisations
- Predicting churns depending on the consumer booking history
- Tracking the real-time weight of the cylinder provides insights on consumer usage patterns

Manufacturers

- Track dealer's performance
- Track service team performance
- Plan out strategies to meet the gas demands with area-wise, season-wise, and dealer-wise projections.

Consumers

- Can track how long a cylinder is lasting.
- Manage bookings when gas levels are low
- Tackling gas leakages
- Can get safety alerts
- The trolley can track environmental information like the temperature to detect any unfavorable conditions.

Data Extraction and Processing Pipelines

Our proprietary architecture built using some of the industry-standard protocols and tools like MQTT, Kafka, NoSQL data stores allows us to set up a data pipeline from the source to the analytical data store with relative ease. This allows us to apply pre-designed learning models to data as it is being generated.

Data Analysis

In addition to pre-designed and pre-trained models, having a continuous data stream that can pick up new data points or variations allows us to refine and tweak the models over time or develop new models altogether. Our expert Data Science team can periodically monitor the efficiency and performance of the models and update them over time

Visualization and Integration Platform

The results of the analysis are delivered through our proprietary Visualization and Integration platform. The platform can be integrated into your Products, Systems, and Processes by Single Sign-On (SSO) or APIs.

Along with features such as continuous monitoring and automated alerts, we can automate an otherwise manual and tedious process such as data analysis and help you build intelligent frameworks from your most valuable resource, your data.

About Qualetics Data Machines

Qualetics Data Machines is a venture-funded startup established in July 2018, in the state of New Jersey in the USA. The company was founded and is led by the CEO, Sumanth Vakada, a software industry veteran with 20-years of experience in building modern software applications across various business domains. Qualetics currently has its core team based in the USA and an extended Application Development and Data Science team based in Hyderabad, India.

Our mission is to enable organizations to make an easier transition into the fields of Data Analytics and Artificial Intelligence. We aim to achieve this more efficiently than an organization might expect, were it to invest in the manpower and technology to build such complex platforms. For a universal need such as Data Science and AI, a dedicated focus is absolutely critical and Qualetics empowers our clients with the knowledge that allows them to apply in their core business offerings.

Key Features

Data Ingestion Platform

Visual Insights Delivery

License Pre-Existing Models

API based Intelligence Delivery

Integration with Apps

Develop Custom Models & Insights

For Inquiries, please contact:

Mike Fowler

Chief Commercial Officer

mike@qualetics.com

630.715.4540

www.qualetics.com