



Data Scientist @ Con Edison | EX- MLE | EX-SDE | 3+ YOE  
Forecasting, Recommendation | I find anomalies to mitigate risk & maximize efficiency



# About Amogh

Hello, I'm an accomplished Analyst with passion for transforming complex datasets into actionable insights and strategic solutions across finance and various sectors. Holding a Master's in Information Systems, my expertise spans advanced statistical analysis, machine learning, AI technologies, and programming with a deep dive into **EXCEL, PYTHON, SQL, POWER BI** and more.

At the forefront of Data Science, I've engineered impactful models and systems, from **Forecasting** and **Recommendations** to optimizing operations and enhancing decision-making processes. My work with Con Edison and Bullwhip Technologies, among others, showcases my ability to leverage data for **predictive analytics, operational improvements, and insightful visualizations**, significantly affecting bottom lines and operational efficiency.

Certified in AWS Cloud, Google Business Intelligence, and as a Meta Database Engineer, I'm not just about data. I thrive in environments that challenge my problem-solving, agile methodologies, and stakeholder management skills. With a track record of leading projects from ideation through deployment, I excel in turning data into a compelling narrative for both technical and non-technical audiences.

Whether streamlining processes, automating data collection, or driving user-centered solutions, I'm dedicated to pushing the boundaries of what data can do to solve real-world problems, making data-driven decision-making the cornerstone of business strategy and innovation.



## Motivation

- The motivation behind this is to empower the creation of high value content on streaming service like YouTube.
- The underlying fundamental is to build a scalable decentralized framework for ingesting data from different sources, training ML algorithms and predicting user content
- Scope is extended to show the effectiveness of Spark framework on large datasets

## Technology

- PYSPARK
- PYTHON
- FLASK
- PANDAS
- HTML/CSS
- JOBLIB
- LIGHTGBM

## Workflow

### Enter Youtube Channel URL

Channel URL

### Enter Youtube Channel URL

Channel URL

### Enter Youtube Channel Title and Description

Channel Title  
  
Channel Description

Classifier Name	Value
GaussianNB	0
LGBMClassifier	1
XGBClassifier	0
AdaBoostClassifier	1

### Word Clouds on all the video data present in the Channel



### ML Algorithms and Analysis Results

Classifier Name	Accuracy (%)	Precision (%)	Recall (%)	F1_Score (%)
GaussianNB	75.87	75.89	75.87	75.86
LGBMClassifier	80.95	80.96	80.95	80.95
XGBClassifier	80.51	80.51	80.51	80.51
AdaBoostClassifier	76.26	76.27	76.26	76.26

### Spark Analysis Results

Classifier Name	Accuracy (%)	Precision (%)	Recall (%)	F1_Score (%)
LogisticRegression	82.36	84.25	93.68	81.15
GBClassifier	84.17	86.42	93.71	83.32
NaiveBayes	80.88	92.23	81.84	81.77

### Run time Analysis

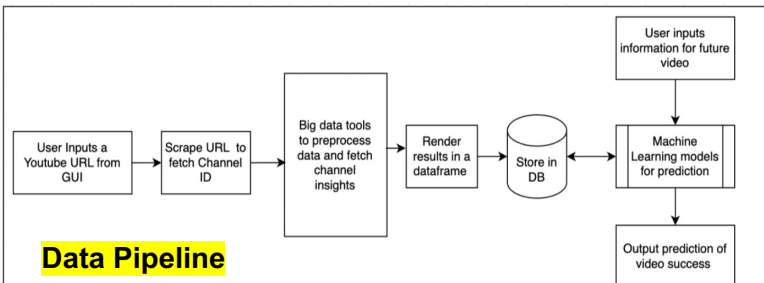
Normal (seconds)	Spark (seconds)
376.81	192.06

### Data shape and class size

Rows x Cols	Rows (0)	Cols (1)
(9055, 3)	4540	4515

Click below to predict how many likes your video might get: [Click Me](#)

## Data Pipeline



# Big Data Text Analysis - YouTube Channel Video Classification

[\[Project Slides\]](#)

[\[DEMO\]](#)

## Conclusion

- The end-to-end analytics as a service based on decentralized microservices can be used in creating a robust real time prediction that can help content creators attract crowd on YouTube
- Although the limitation to the current development is on untuned algorithm but they tend to perform well in binary classification (0,1), auto tuning can be applied to make them better
- we believe the proposed big data architecture will be widely extensible and expandable that can be produced as industry grade service for everyone

## Current & Future Work

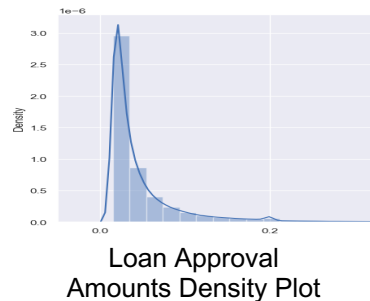
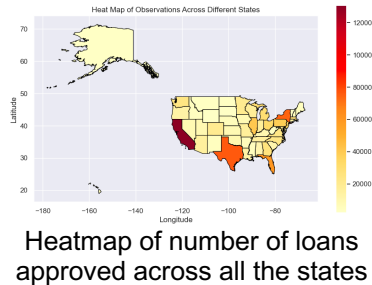
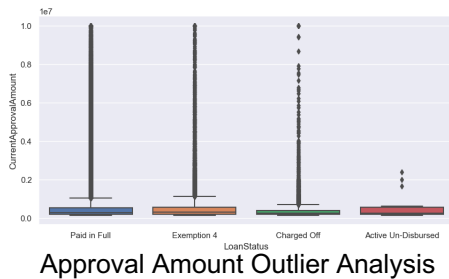
- The current scope consists of analyzing data from one channel from YouTube to train on ML algorithms and Spark based to create predictions from user and display result on Frontend using Flask
- The current scope of deriving insights from one channel can be expanded to multiple channel input
- The preprocessing and ML modeling can be parallelized by running multiple instances
- Decentralization and guarding it from short term failure can be resolved using AWS ecosystem the entire project can be cloud based to handle high volume of requests and process them in real time paving the way of real time streaming analytics and prediction



# Fraud Detection - Paycheck Protection Program

## Exploratory Analysis

- Investigated data to understand distribution, anomaly with heat map and charts
  - Others were specific to data exploration with Python
- Some examples of this effort can be seen below.



## Motivation

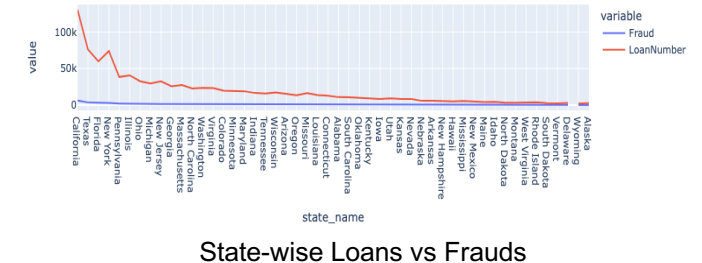
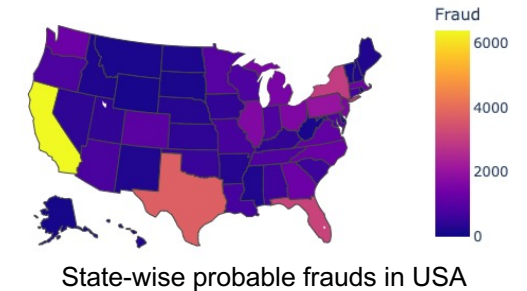
- Explore loan data from the Paycheck Protection Program (PPP), which provided relief to small and medium-sized businesses during the COVID-19 pandemic and identify probable fraudulent loans.
- The primary objective is to reduce frauds in the future by applying anomaly detection methods to identify outliers and building machine learning models to potentially detect possible frauds.

## Anomaly Detection

- Data is Standardization using standard scalar
- Findings Higher the anomaly score, higher the probability of loan being fraud
- Used Isolation forest to calculate anomaly scores
- Calculated Anomaly score for features:
  - Initial Approval Amount
  - Payroll proceed
  - Jobs Reported
- Identified loan being fraud biases of anomaly score top 95% = anomaly (i.e. fraud)

## Technology

- Python was used for data pre-processing and cleaning
- Python libraries matplotlib, plotly, seaborn utilized for exploratory analysis and dashboarding
- Python to identify anomalies using pandas, matplotlib, IsolationForest in Sklearn



## Current & Future Work

- Government agencies and policymakers may use the data to evaluate the effectiveness of the program and make necessary adjustments to future relief efforts
- Researchers may use data to study the economic impact of the pandemic on small businesses and to identify patterns and trends in the distribution of PPP loans
- Incorporate possible frauds with portal like Datamerch, Experian, lapps to improve due diligence for debt-based Venture Capitalists

## Goal Based Student Diet Personalization

(Interviewed students using **Design Thinking** tools (Empathy), analyzed data to understand user needs to develop product to save busy student's deteriorating health)

[\[Prototype\]](#)

[\[Project Slides\]](#)

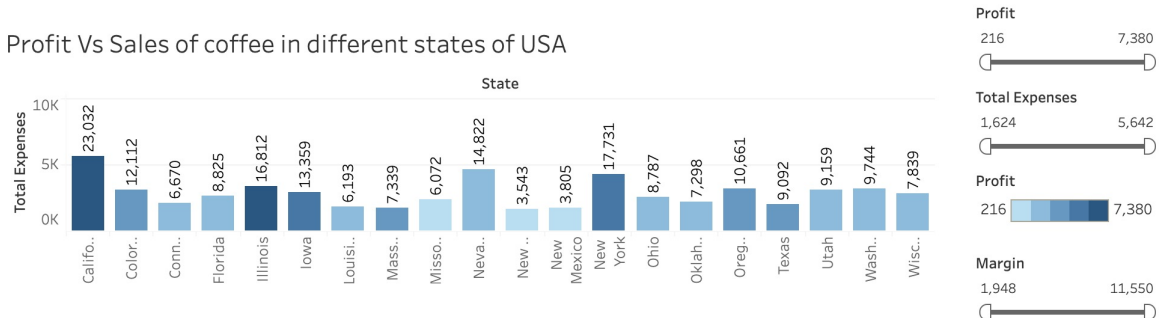
[\[Explainer Video\]](#)



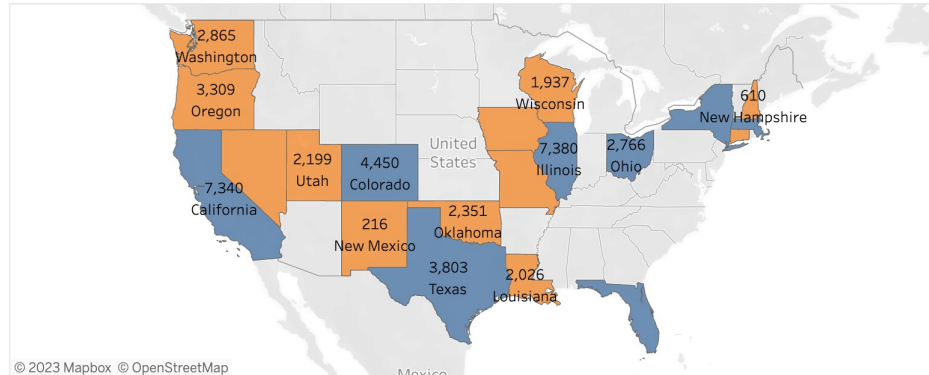
# Tableau Viz on Profit VS Sales of Coffee in Different States of USA

[LINK]

Profit Vs Sales of coffee in different states of USA

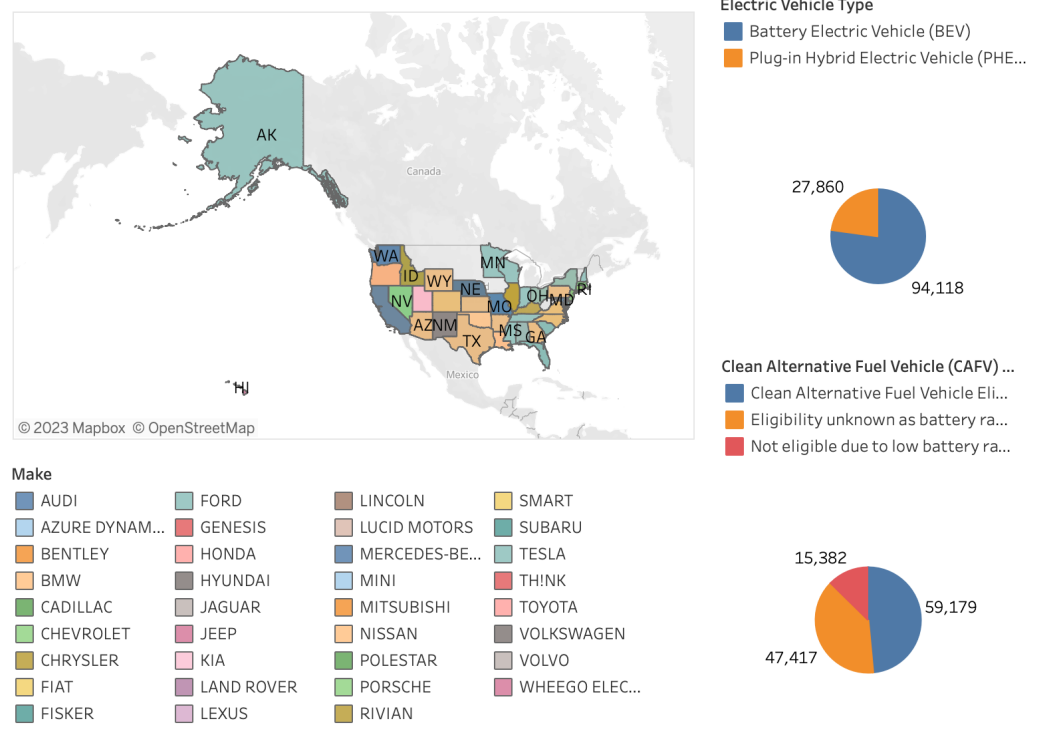


Map



# Electric Vehicle Analysis [LINK]

Electric Vehicle Population Data [Washington US]



## Key Accomplishments

**Power Grid Optimization at Con Edison:** Mapped 1M+ NYC grid components, developing a machine learning model that predicted EV charging with 96% accuracy, optimizing charging times to reduce grid stress and cut electricity bills.

**Keyword Strategy Innovation at Bullwhip Technologies:** Created a time series forecasting model and a recommendation system, boosting keyword prediction accuracy and increasing user engagement and revenue by 15%.

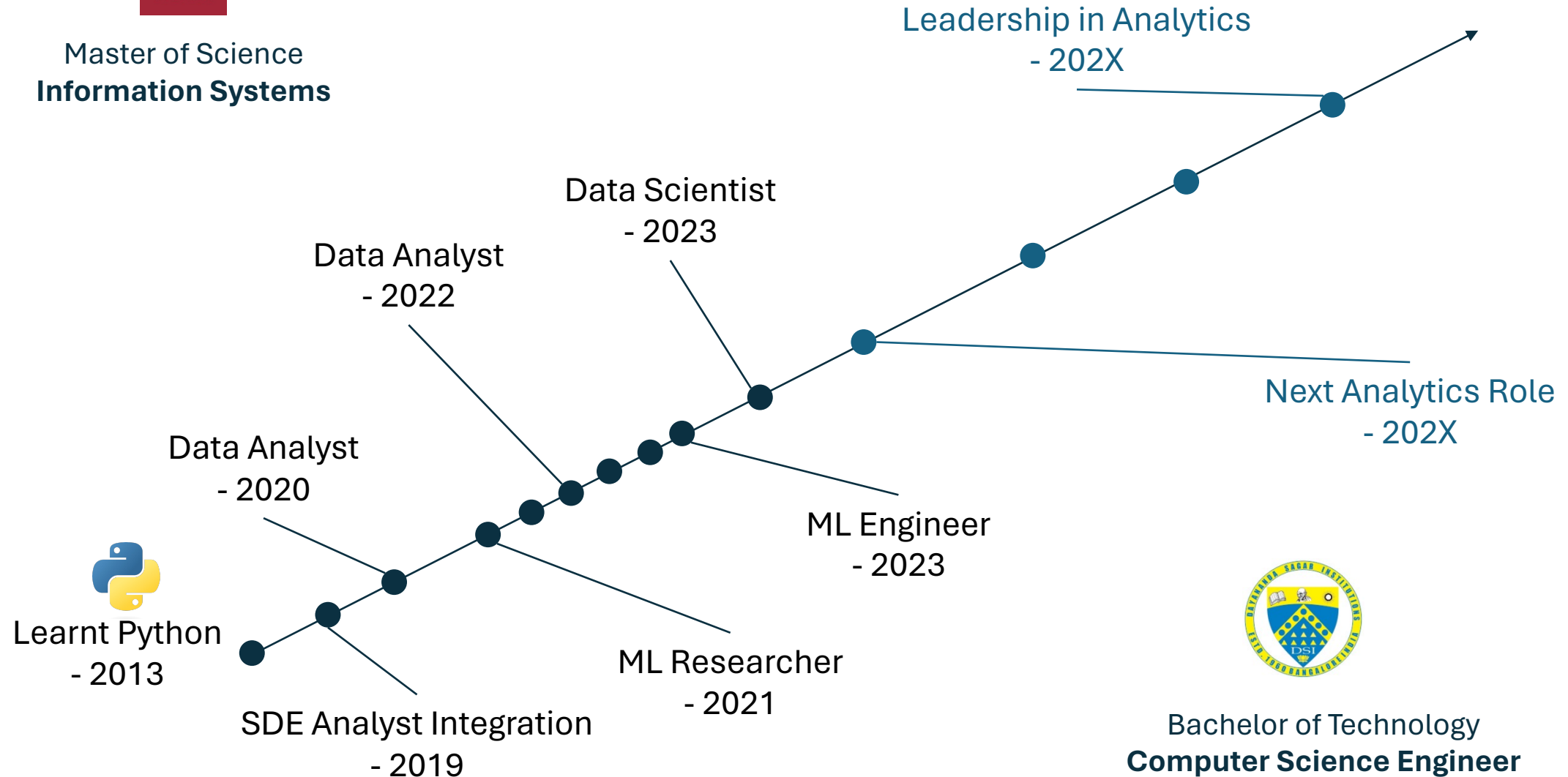
**Operational Efficiency Boost at Build Health International:** Developed a financial oversight dashboard, significantly streamlining project management and reducing employee transition time from one day to one hour.

**Payment Security Enhancement at Razorpay, Nioneer:** Led fraud detection on 1M+ transactions and automated payment processes, saving \$60,000 monthly, while improving digital engagement by 20% through targeted marketing strategies.



Master of Science  
**Information Systems**

## CAREER TIMELINE





# Testimonials

---

**Chinmay Rathod** (Co-Founder, Lead Data Scientist)

I had the pleasure of working with Amogh on a beta testing project that involved deploying code on AWS. His strong knowledge of cloud computing, software development, and data science was incredibly helpful for the project .....

---

---

**Balaji Rao** (PhD, Blockchain Researcher)

I worked closely with Amogh for over a year as a research assistant at the Stevens Institute of Technology. Amogh is a highly skilled individual with expertise in software development, data engineering, and data science. He built the Human-Computer Interaction Lab's official website, collaborating with a team of 5 researchers and developers to create an interactive user experience .....

---

## References available on request

**Michael Washington**  
Director of Finance

**Dr Thomas Lechlar**  
Project Management Professional

**Dr Edward Stohr**  
BIA Program Director

**RJ Lehman**  
Founder Stealth Startup

**Dr Joseph Morabito**  
Business Intelligence Professional

# Thank You

[LINKEDIN.COM/IN/AMOGHKOKARI/](https://www.linkedin.com/in/amoghkokari/)

[GITHUB.COM/AMOGHKOKARI](https://github.com/amoghkokari)

[GETFOODRECIPE.STREAMLIT.APP](https://getfoodrecipe.streamlit.app)

