Profile Overview

Seasoned Data Scientist with 11 years of experience in enabling data led decision making. Passionate about telling stories from data. Keen to utilize my analytics solution building exposure in enabling business transformation

Profile:

Name: Utkarsh Tripathi

Organization: Micron Technology

Location: Hyderabad

Designation: Principal Data Scientist

Experience: 11+ years

Education:

Bachelor of Engineering, ECE, IPS Academy

PG Diploma in AI & ML from IIIT-Bangalore

MS in AI & ML from Liverpool John Moore's University

Organizations Worked:

Robert Bosch: Oct 2014 – Feb 2018

Times Internet: Mar 2018 – Nov 2018

Publicis Sapient: Dec 2018 – May 2020

Fidelity International: May 2020 – April 2021

Micron Technology: April 2021- Present

Chegg Inc. - Subject Matter Expert Statistics and Probability (part-time):

April 2018 – Dec 2019

Profile Link

LinkedIn: **https://www.linkedin.com/in/utkarsh-tripathi/**

GitHub: **https://github.com/utkarsh1993**

Agenda

- **Achievements and Extra-Curricular Activities**

- **AI & ML Projects**

  o **Cost 360 Assistant**

  o **Text2SQL using RAG**

  o **Attrition Predictor**

  o **Ad Unit Segmentation**

Achievements and Extra-Curricular Activities

Achievements

Extra Curricular Activities

- Published paper titled "Database Search: Text2SQL using dynamic few-shot prompting with self-consistency using LLM" in Journal Lattice and MLDS 2025 conference organized by AIM

- Led the team and won 1st Prize in hackathon Dev Premier League – Organized by Hack2Skill and Snowflake

- Secured Place in Top 91 (Top 1%) teams in Google Agentic AI day – presented by Google and Hack2skill

- Second Runner Up In Howathon and Sapient Expo India 2019

- Kaggle Competitions Expert: 2 Bronze and 1 Silver medal
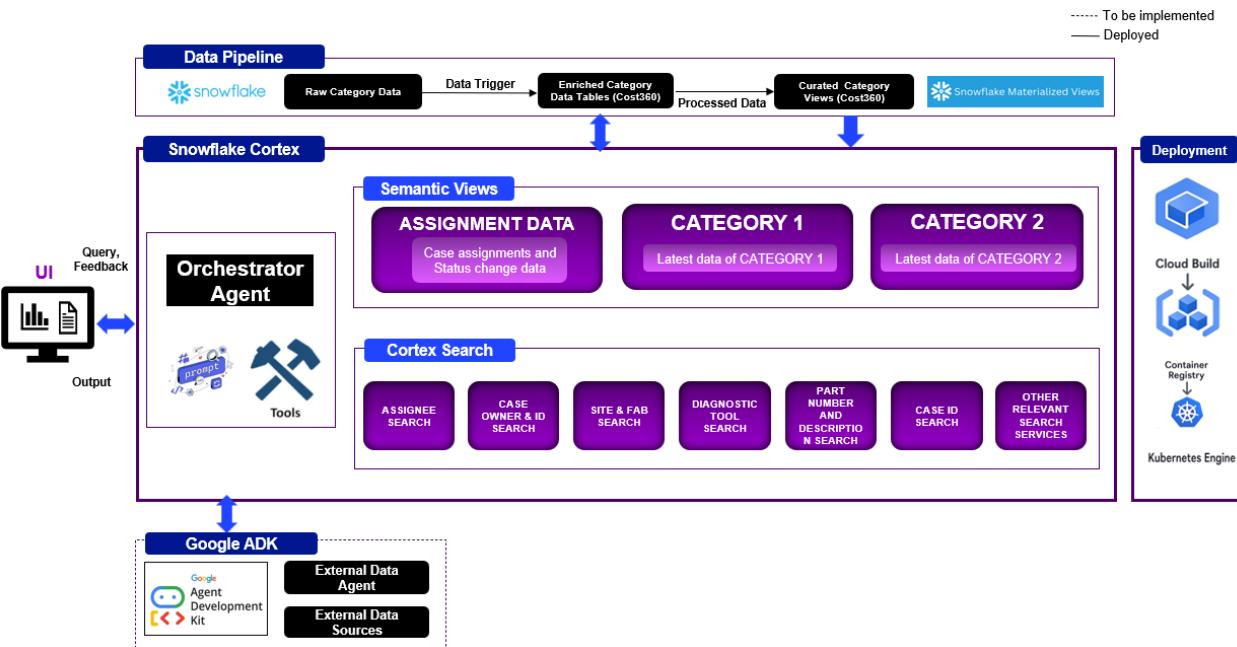
- Multiple certifications from Coursera on AI, Machine Learning and Deep Learning

- Participated in Gen AI Exchange Hackathon

- Oracle Certified Java Programmer –Professional SE6
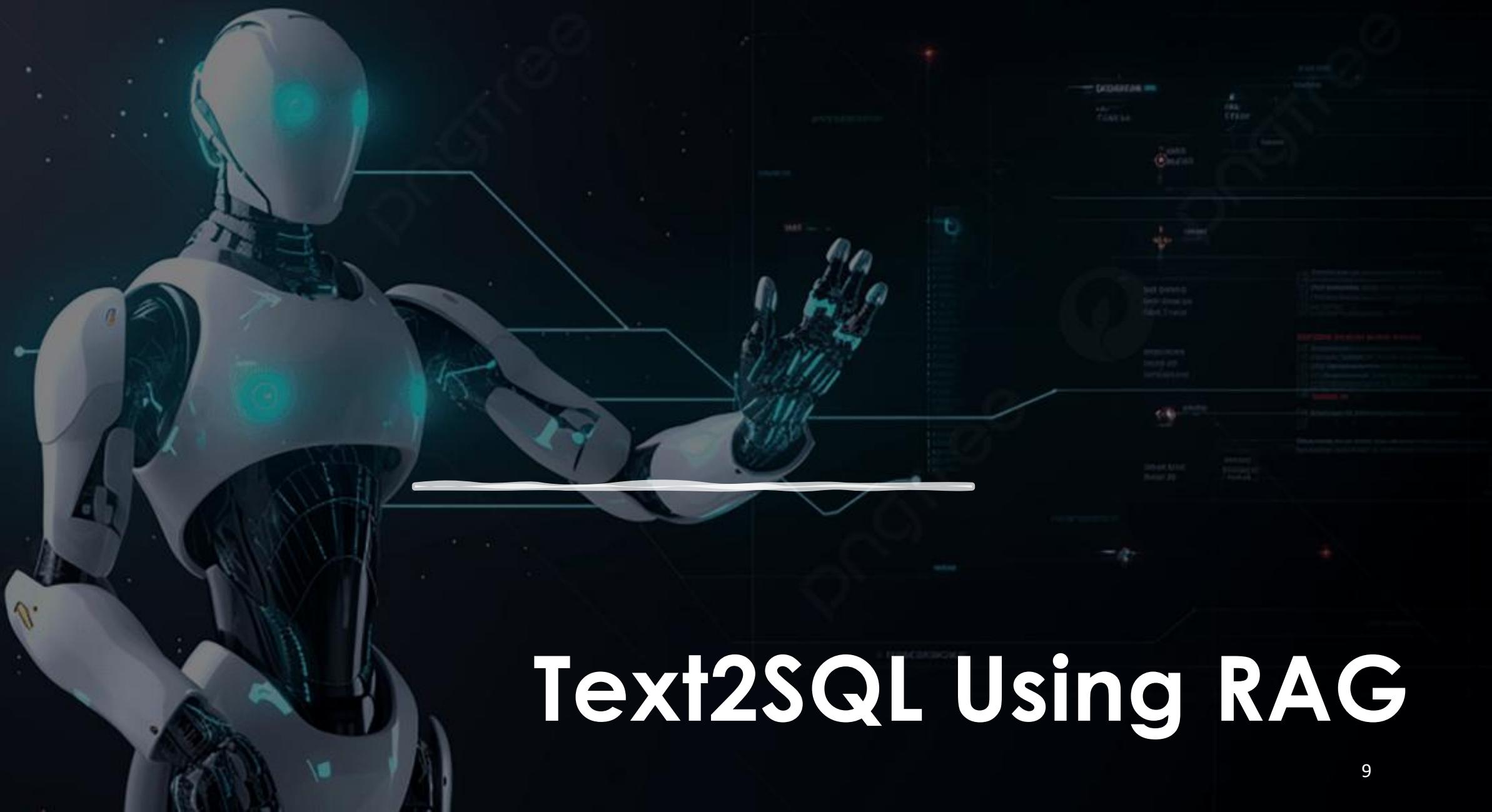
Cost360 Assistant

High Level Architecture

Cost360 Assistant leverages Snowflake Cortex and Agentic AI to streamline procurement cost management. Cortex Agents autonomously gather and process data, enabling natural language updates and real-time insights. Combined with Cortex Analyst and Search Service, this system simplifies workflows, accelerates analysis, and empowers procurement teams with actionable intelligence and seamless control.
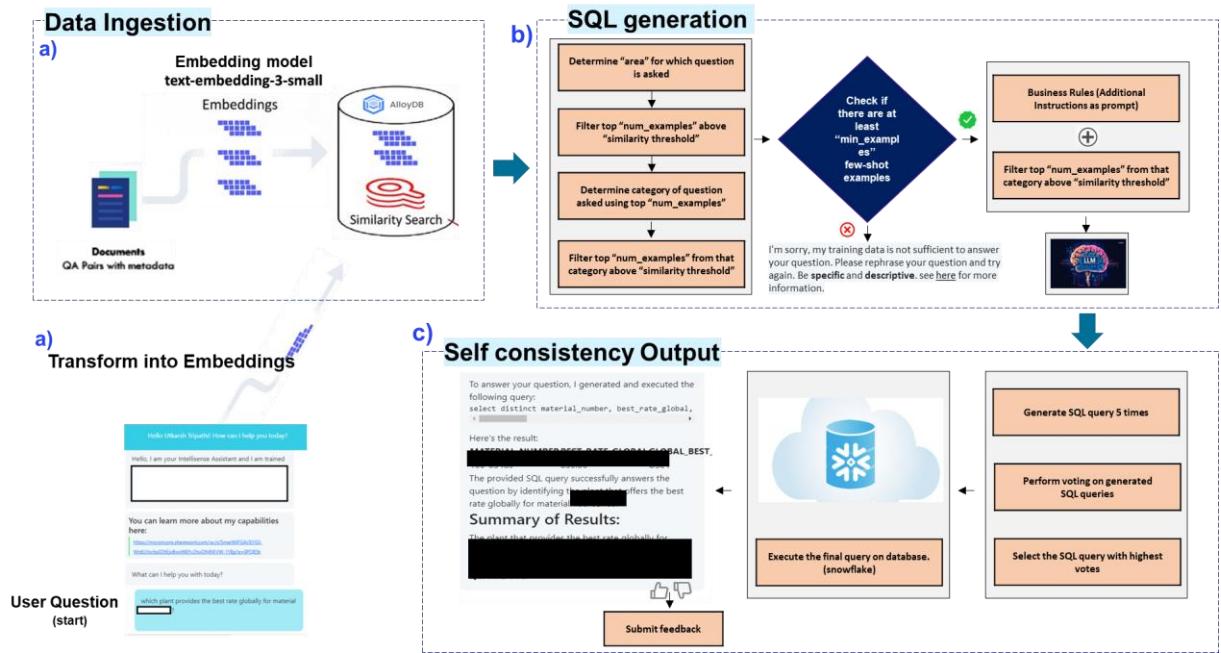

Activity	Before Cost360	After Cost360	Improvement
Update 100 Case Assignments	~3 hours (manual updates)	< 2 minutes (bulk update)	>99% faster
Identify Savings Opportunities	~4–6 hours (manual data gathering)	< 10 minutes (case summary)	>95% faster
Time Spent on Manual Tasks	30–40% of working hours	<10% of working hours	30–40% productivity gain

Text2SQL Using RAG

High Level Architecture

Text2SQL enables non-experts to query relational databases using natural language. Traditional fine-tuning risks overfitting, while LLM-based prompting improves adaptability. Our dynamic RAG framework scales across multiple tables using few-shot examples, business rules, and self-consistency voting. It integrates seamlessly into analytical platforms, delivering accurate SQL queries, results, and summaries for faster decision-making.

Area	Category	Number of Questions	Correct Answers	Not Enough Data for training*	Wrong Answer	Accuracy	Overall Accuracy
proc/category-x/sol-1	Spend	176	160	14	2	90.9%	93.39%
	Negotiaion	190	181	7	2	95.3%	
	Clarification	300	281	18	1	93.7%	
proc/category-y/sol-2	Spend	57	55	2	0	96.5%	95.97%
	Opportunity	65	60	0	5	92.3%	
	Arbitrage opportunity	150	144	0	6	96.0%	
	Trend Opportunity	150	146	0	4	97.3%	

Attrition Predictor

High Level Architecture

Objective

To create a system which can improve on the attrition rate by
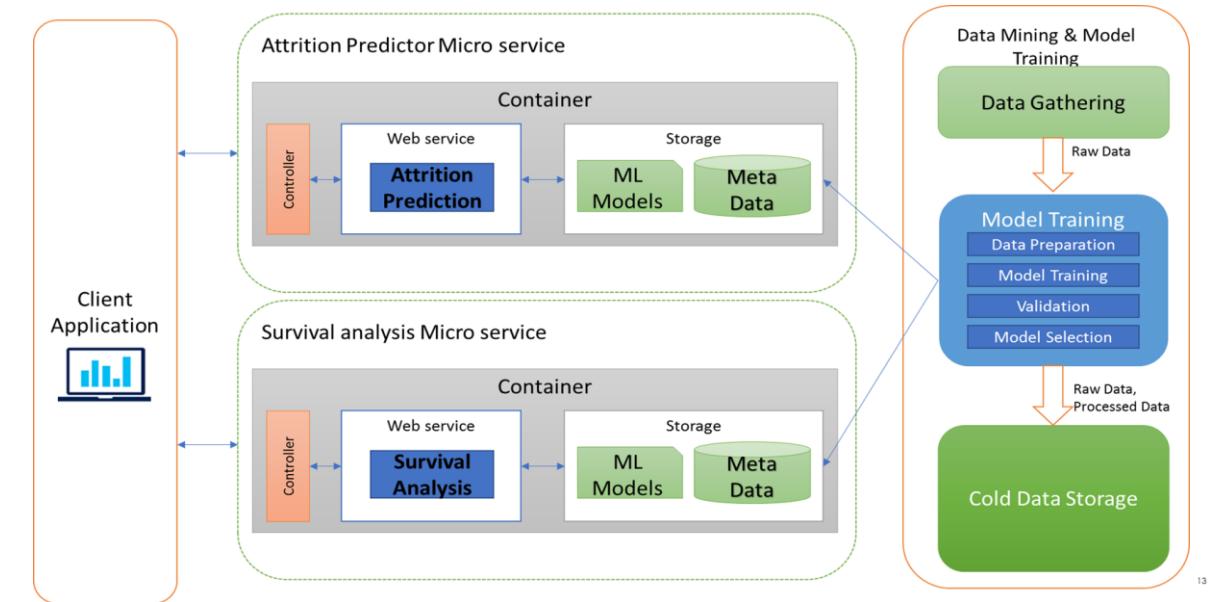
- Predicting the probability that an employee will leave
- Suggest the most probable reasons for an employee leaving the organization
- Identify the survival rate of employees in the organization

Solution Implemented

Attrition Predictor is a simple web app to identify the probability of an employee leaving the company along with the most likely reasons.
Web application has following features:
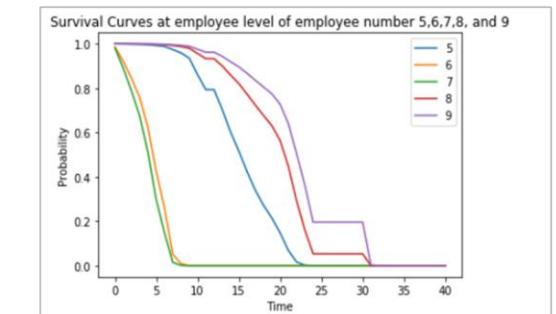
- Dashboard showing current attrition rate and the top 5 contributing factors for attrition
- Upload a new file with employee data to get attrition probability for each employee along with a probable reason,
- Upload a new file with employee data to get the survival analysis outcome for 0.5, 1 and 2 year
- Upload historical data to retrain the model
- Framework reusable for attrition prediction on different data sets & customer churn analysis

Architecture diagram

Client Application

Attrition Predictor Micro service

Container
- Web service — Attrition Prediction
- Storage — ML Models, Meta Data
- Controller

Survival analysis Micro service

Container
- Web service — Survival Analysis
- Storage — ML Models, Meta Data
- Controller

Data Mining & Model Training
- Data Gathering
- Raw Data
- Model Training
  - Data Preparation
  - Model Training
  - Validation
  - Model Selection
- Raw Data, Processed Data
- Cold Data Storage

13

Results

1. 75 to 80% Accuracy on predicting the churn of Employee along with the most probable reason.

2. Important Variables based on their Score (Gini)

	Score
Age	0.110760
MonthlyIncome	0.087104
DailyRate	0.071556
DistanceFromHome	0.068564
MonthlyRate	0.066299
HourlyRate	0.061126
PercentSalaryHike	0.046278
NumCompaniesWorked	0.042979
TotalWorkingYears	0.037750
over_time_2	0.037358
YearsSinceLastPromotion	0.034366
Env_satisfaction_3	0.033080
YearsAtCompany	0.032549
Department_new_2	0.024859

	coef	exp(coef)	se(coef)	z	p	-log2(p)	lower 0.95	upper 0.95
Age	-0.17	0.85	0.10	-1.69	0.09	3.46	-0.36	0.03
DailyRate	-0.03	0.97	0.07	-0.50	0.62	0.70	-0.17	0.10
DistanceFromHome	0.21	1.24	0.07	3.20	<0.005	9.51	0.08	0.34
Education	0.01	1.01	0.07	0.14	0.89	0.17	-0.13	0.15
EmployeeNumber	-0.06	0.94	0.07	-0.89	0.38	1.41	-0.20	0.08
EnvironmentSatisfaction	-0.34	0.71	0.07	-4.90	<0.005	19.98	-0.47	-0.20
HourlyRate	0.04	1.04	0.07	0.54	0.59	0.77	-0.10	0.18
JobInvolvement	-0.18	0.83	0.06	-2.82	<0.005	7.69	-0.31	-0.06
JobLevel	-0.09	0.92	0.30	-0.29	0.77	0.37	-0.67	0.50
JobSatisfaction	-0.40	0.67	0.07	-5.78	<0.005	26.97	-0.54	-0.27
MonthlyIncome	-0.16	0.85	0.31	-0.50	0.62	0.70	-0.77	0.46
MonthlyRate	0.11	1.11	0.07	1.52	0.13	2.96	-0.03	0.25
NumCompaniesWorked	0.63	1.87	0.07	8.51	<0.005	55.72	0.48	0.77

Survival Curves at employee level of employee number 5,6,7,8, and 9

Ad Unit Segmentation
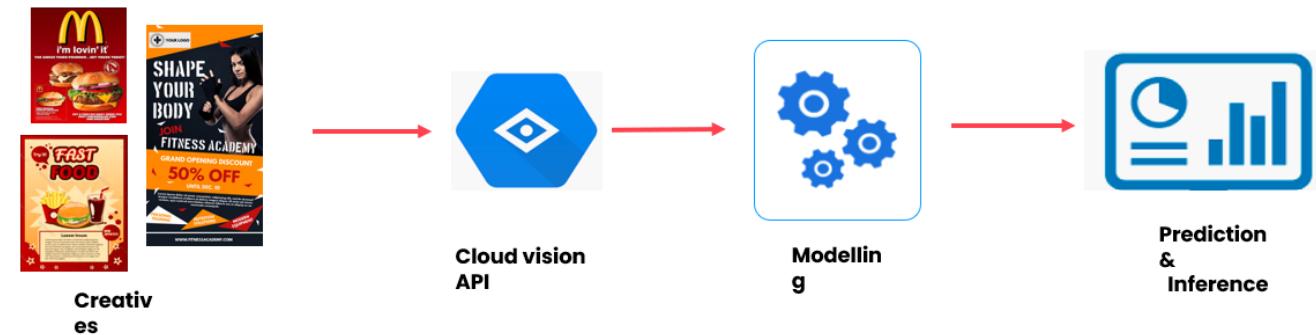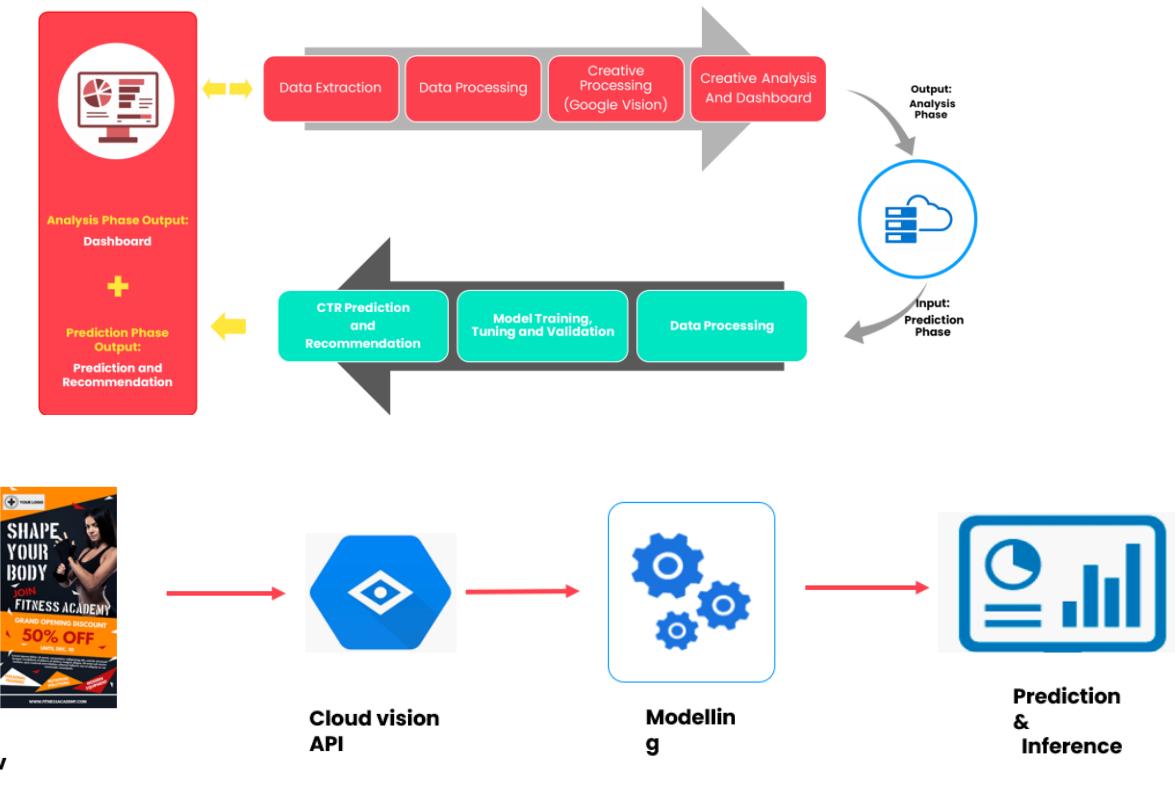
High Level Architecture

Objective

Objective was to group various Ad units-based present on different Web sites and Mobile Apps based on their properties

Solution Implemented

- Around 0.1 million Ad units are there from different products like Mobile App, Website and FB Pages

- Derived new features like Position of Ad units, Cost per million etc.

- Later, Clustering was performed for Ad units generating Revenue of more than 0.1 million INR in 3 months

- Clustered Ad units helped in defining Strategy and pricing for each group

Observation	Key Features	Impact on CTR	Statistical Test Results
On text analysis it is evident that few keywords are having positive impact on CTR	WIN OFFERS PRIZES	4x vs 0.38% mean	Significance testing proves that at 95% confidence interval there is an impact of keywords (Win/Offers/Prizes) on CTR
Prominent colors used in creatives are impacting the performance	Brand colors Non Brand colors	Brand :2x vs 0.38% mean Non Brand :0.5x vs 0.38% mean	Significance testing proves that at 95% confidence interval, brand colors will have an impact on CTR
Creatives used for stress targeting has shown positive impact on CTR	For a limited time	2x vs 0.38% mean	Significance testing proves that at 95% confidence interval, limited time offers will have an impact on CTR
Text length has also shown an impact on the performance of CTR	<20 20-50 51-80 >80	<20 : 0.25x vs 0.38% mean 20-50 : 0.5x vs 0.38% mean 51-80 : Same as mean >80 : 5x vs 0.38% mean	Data is too less

Thank You