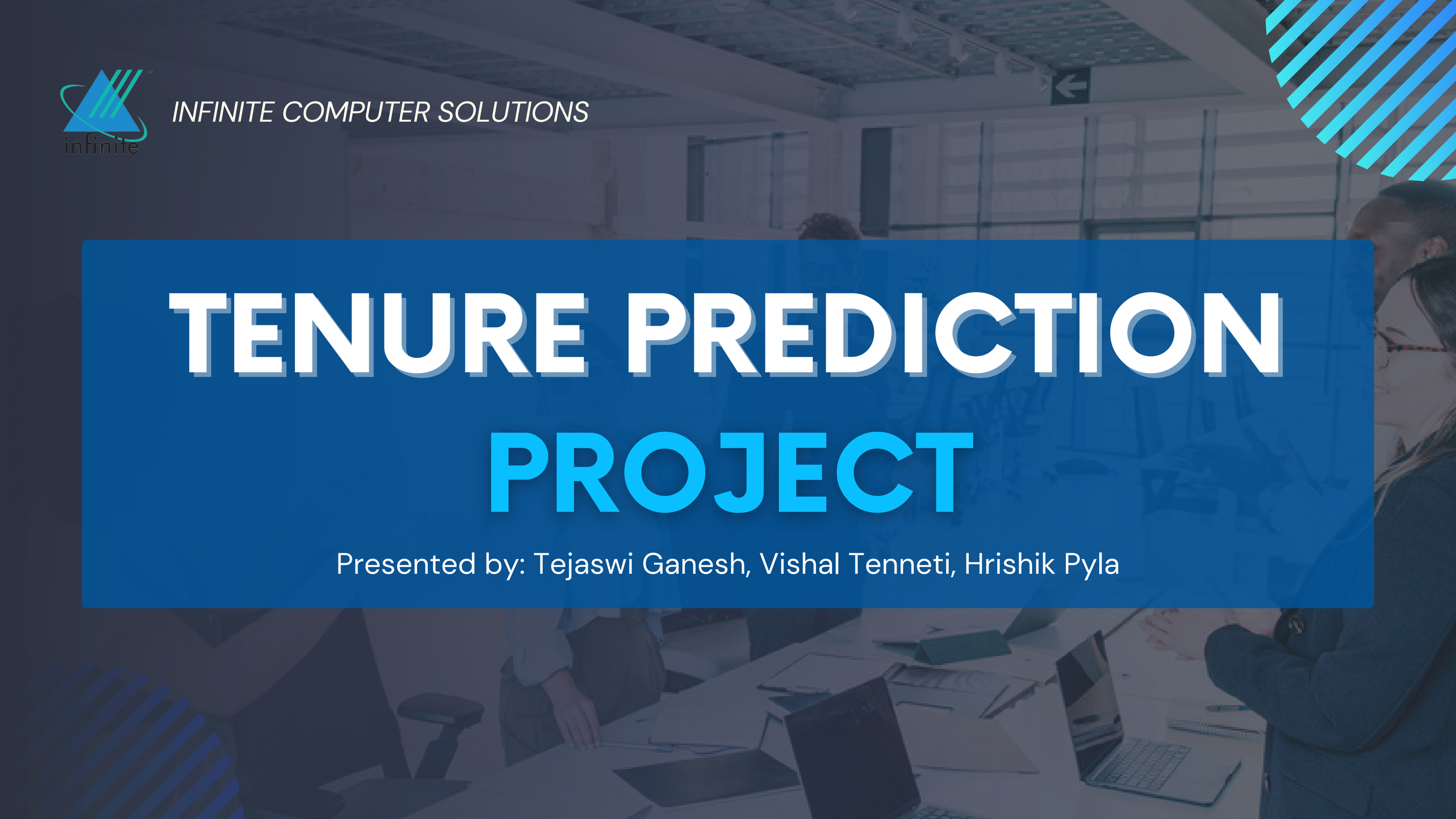


INFINITE COMPUTER SOLUTIONS

TENURE PREDICTION PROJECT

Presented by: Tejaswi Ganesh, Vishal Tenneti, Hrishik Pyla

OVERVIEW

- Problem Statement
- Importance Of Data Cleaning
- JobTitle Standardization
- Tiering Of Organization Names
- Importance And Creating DataSets
- Types Of Models
- Training And Evaluation of Model
- Deploying the Model

PROBLEM STATEMENT

To Predict The Number of Years
A New Applicant
Will last in the Company
(if Recruited)
Based on the Past Experience data
received from the Resumes
that the Applicants Submitted

IMPORTANCE OF DATA CLEANING

- JobTitles Can be entered In Any Format By the Applicants

Website Designer

&

Web Application Designer

- All these Titles are to be grouped into a fixed Data consisting of Organised JobTitles

JobTitle Standardisation

Fixed DataSet
Of all the JobTitles

onetonline.org

O*NET-SOC Code	Title	Alternate Title
11-3021.00	Computer and Information Systems Managers	Technical Services Manager
11-3021.00	Computer and Information Systems Managers	Technical Solutions Director
11-3021.00	Computer and Information Systems Managers	Technical Support Content Manager
11-3021.00	Computer and Information Systems Managers	Technology Director
11-3031.00	Financial Managers	Accountant Supervisor
11-3031.00	Financial Managers	Accounting Director

Even If We have a Fixed DataSet

How Did We Go Forward With Classifying Our Titles Into These ONET_Codes

PROBLEMS Faced With O*NET Codes

ONET Code	Title	Alternate Title	Industries	Alt_Industry
11-2022.00	Sales Managers	Business Development Executive	Wholesale Trade (20%), Retail Trade (17'	
11-1011.00	Chief Executives	Business Development Executive	Professional, Scientific, and Technical Se	IT

- **Many Titles got a unique ONET Code By Applying the Industry while Standardising; But There were some with Duplicates Still**

ONET Code	Title	JobTitle	Industries	Alt_Industry
15-1243.01	Data Warehousing Specialists	Analytics Manager	Professional, Scientific, and Technical Service...	IT
15-1243.00	Database Architects	Analytics Manager	Professional, Scientific, and Technical Service...	IT
15-1252.00	Software Developers	Application Integrator	Professional, Scientific, and Technical Service...	IT
15-1253.00	Software Quality Assurance Analysts and Testers	Application Integrator	Professional, Scientific, and Technical Service...	IT

- **These Type of Matches are currently not being considered for Training, But Instead are kept aside in a Table ('Conflicting')**

PROJECT STEPS

Step 1

Job Title Standardization

Step 2

Company Tiering

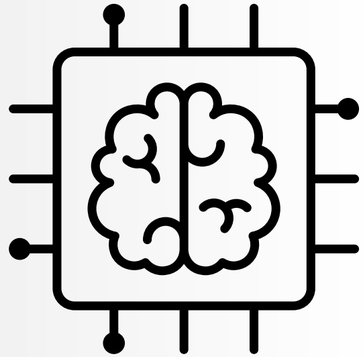
Step 3

Model Training & Deployment

COMPANY TIERING

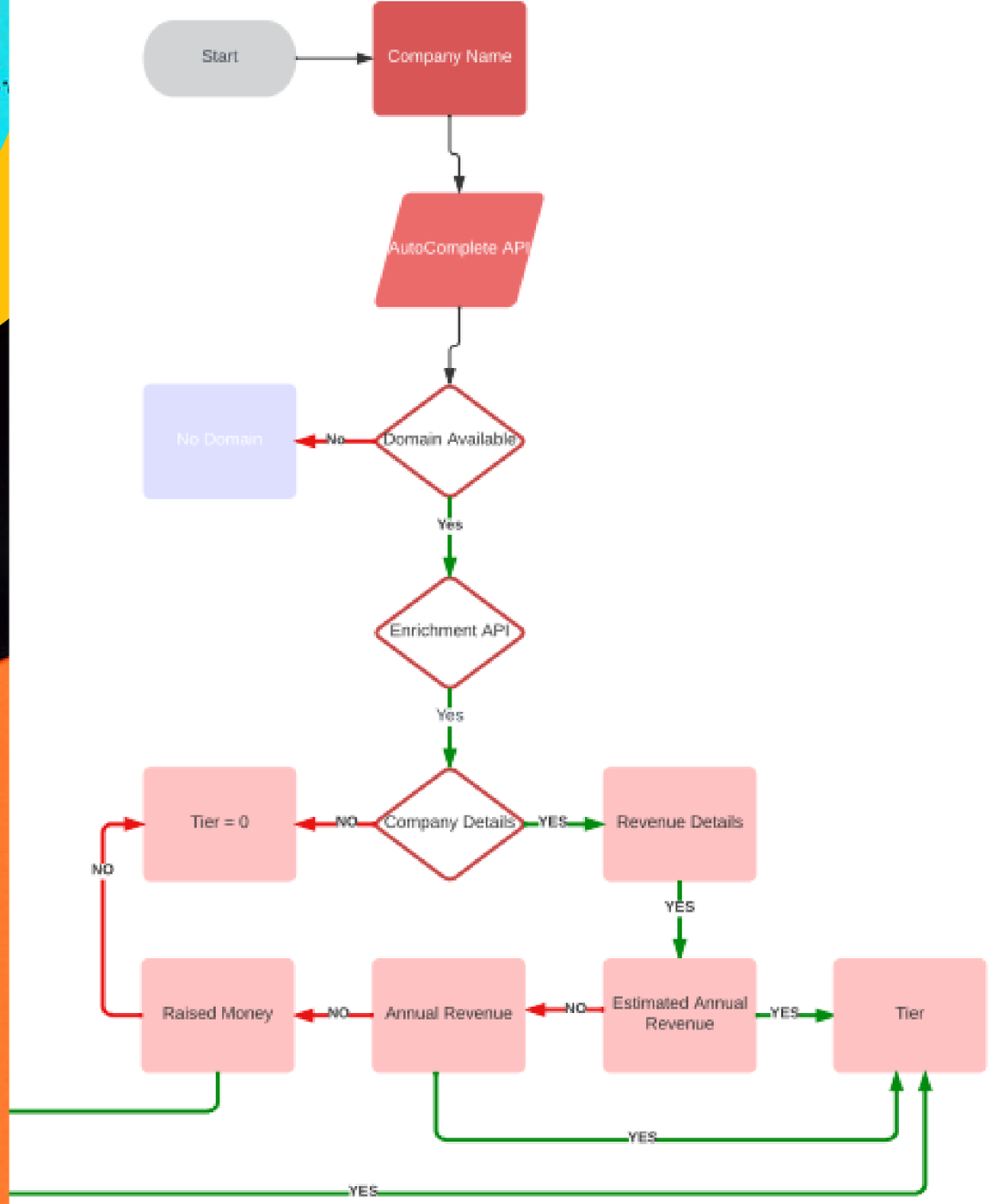
5-Step Process

1. Company Name Standardization
2. AutoComplete API
3. Enrichment API
4. Company Tier
5. Recreate Experience Table



COMPANY NAME STANDARDIZATION

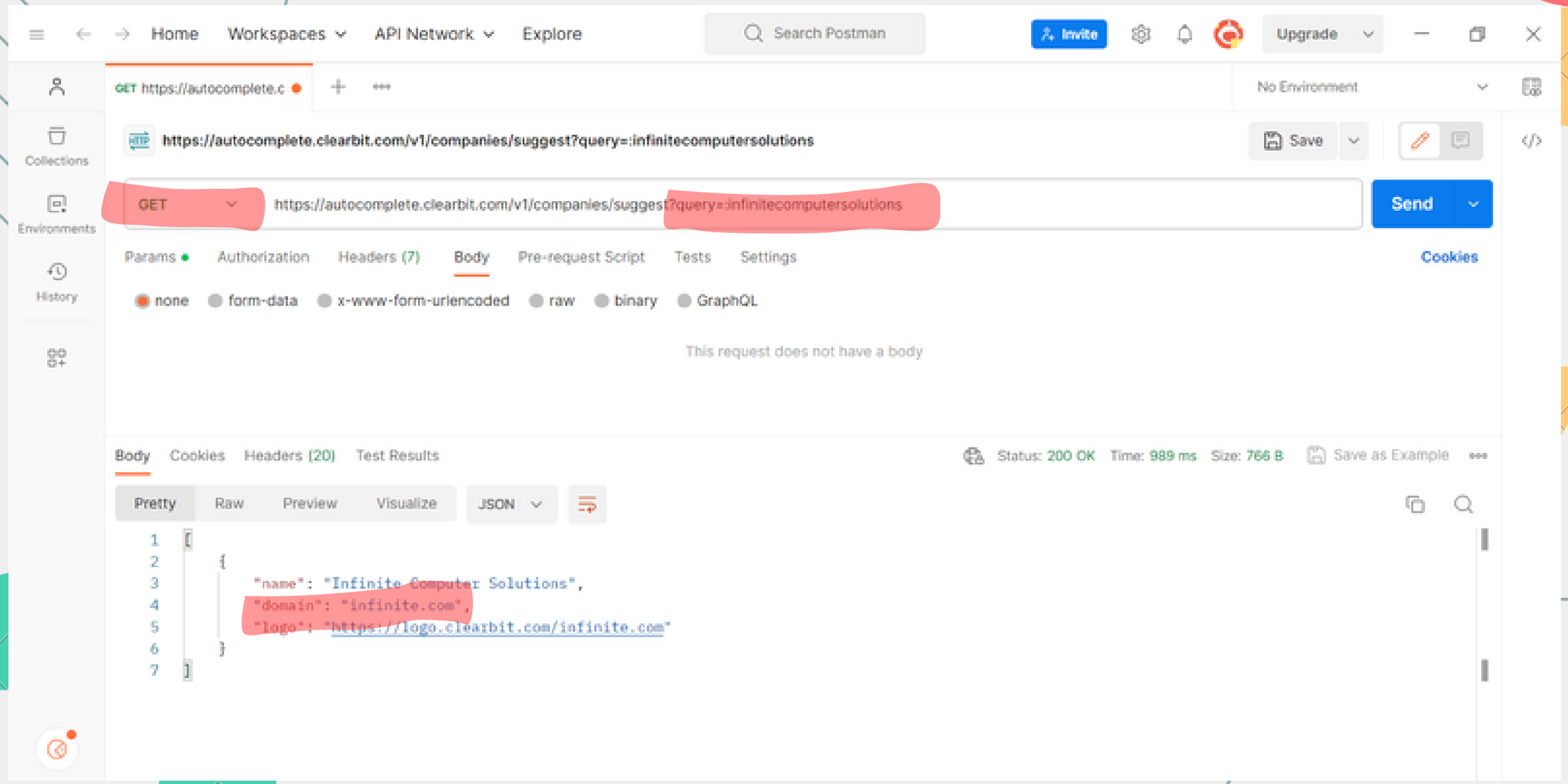
3	Accenture - São Paulo / SP
4	Accenture Services (P) Ltd
5	Accenture – St. Louis, MO
6	Accenture - T&M Consultoria em Informática SAP
7	ACCENTURE - TECHNOLOGY LABS
8	Accenture - Texas Instruments
9	Accenture - US Banking Client
0	Accenture - US Telecomm Client
1	Accenture - Vale
2	Accenture - WHIRLPOOL
3	Accenture - Yesler
4	Accenture & EMC



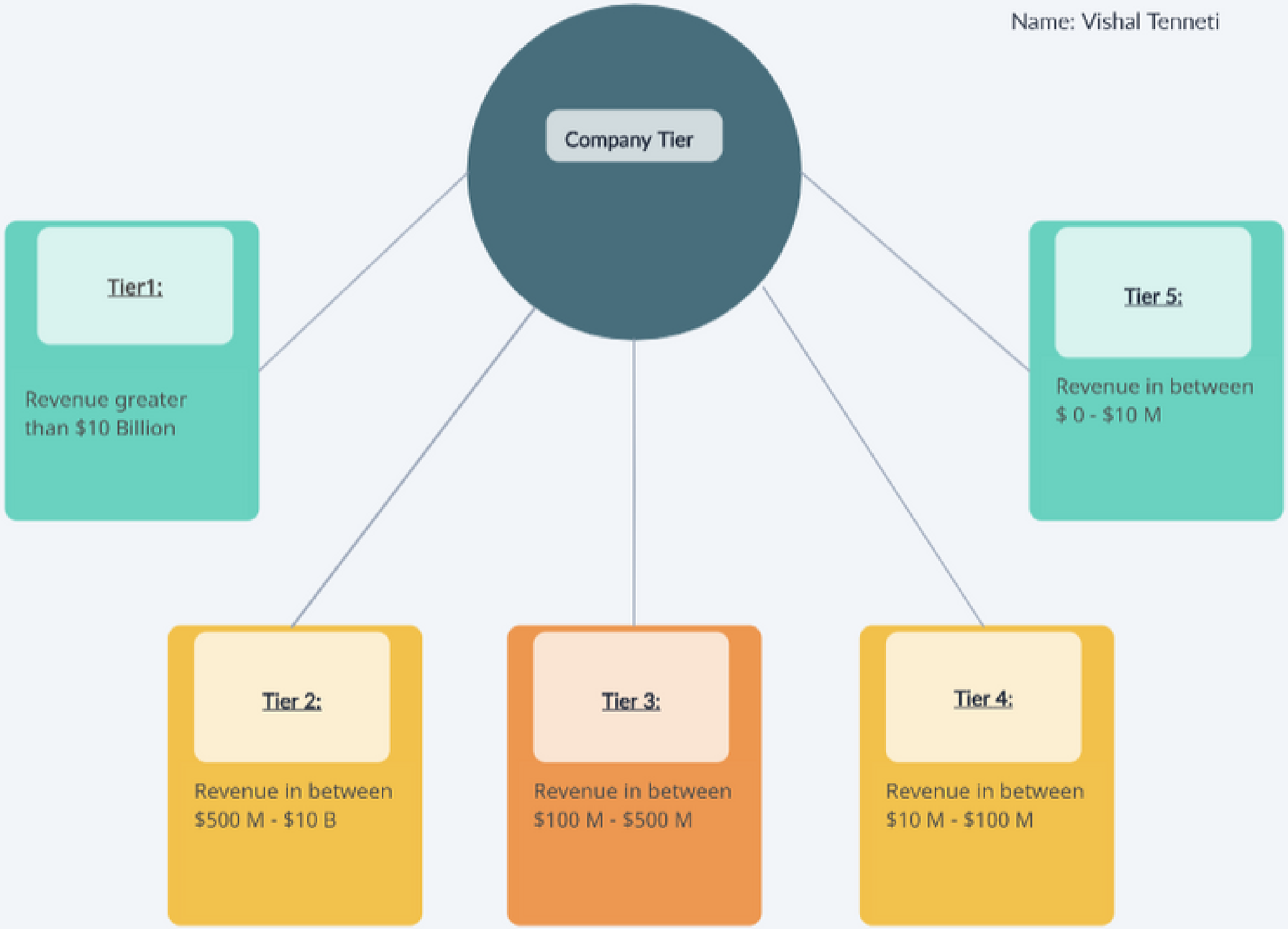
FLOWCHART OF COMPANY TIERING

This diagram provides a clear idea of the process in which a Company is assigned a Tier

AUTOCOMPLETE API



Name: Vishal Tenneti



DATASETS

ONET-CODES

JOB ROLE + TIER

ORG. NAMES

ONET-CODES

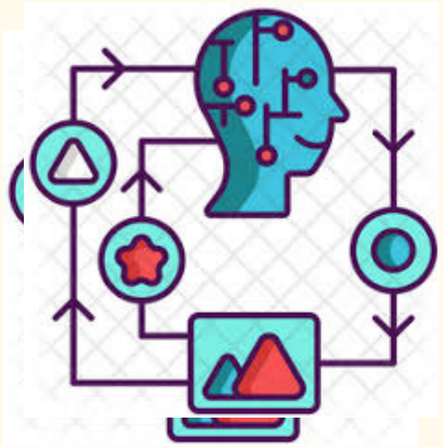
15-1251.00

O H E

ONE HOT ENCODING

	OrgName1	Tier1	Role1	ONET_Code1	Tenure1
1	AMDOCS	2	ORACLE DATABASE ADMINISTRATOR	15-1242.00	2
2	ADROIT Infotech	5	SAP ABAP PROGRAMMER	15-1251.00	3
3	Accenture	1	Software Engineer	15-1252.00	4
4	ADT	2	SURVEY ENGINEER	17-1022.00	1
5	Accenture	1	Software Engineer	15-1252.00	4
6	Accenture	1	Senior Software Engineer	15-1252.00	11
7	Anadolu University	2	Project Assistant	43-6011.00	1
8	Accenture	1	BUSINESS ANALYST	15-2051.01	9
9	Adecco	2	Sales processes support specialist	15-1232.00	4
10	Accenture	1	Quality Assurance Tester	15-1299.04	1
11	Alcatel-Lucent	2	International Technology Consultant	15-1299.00	4

SOME TYPES OF MODELS



Decision Tree

Random Forest

K - Nearest Neighbours

Regression

FOR DATASET-1

X : X1 X2

ONET -CODE1 TIER1

Y1 ⟶ TENURE1

DEPENDENT

FOR DATASET-2

X : X1 , X2 , Y1 X3 X4

ONET-CODE2 TIER2

Y2 ⟶ TENURE2

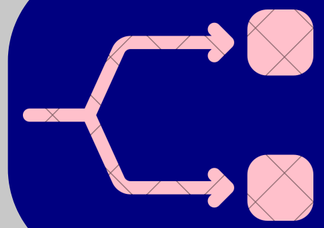
So, Here Y is changing

WHY REGRESSION ?

MODEL

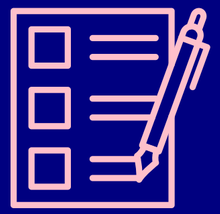
MULTI - LINEAR REGRESSION

$$y = AX1 + BX2 + C$$



SPLIT

TRAIN



TEST

DEPLOY

MODEL EVALUATION WHEN TRAINING DATA IS LESS

```
Number of Rows for training in DataSet 1  =  1248
Number of Rows for testing in DataSet 1  =  313
number of valid predictions =  292
number of inValid predictions =  21
number of inValid predictions which are in trained =  0
R2 Score =  -0.07183052840098059
MSE =  17.646624155240517
MAE =  2.4328044473458904
```

MODEL EVALUATION WHEN TRAINING DATA IS MORE BY GIVING RANDOM TIER TO REMAINING DATA

```
Number of Rows for training in DataSet 1  =  22576
Number of Rows for testing in DataSet 1  =  5645
number of valid predictions =  5611
number of inValid predictions =  34
number of inValid predictions which are in trained =  0
R2 Score =  0.017692614640011O4
MSE =  7.679635533078047
MAE =  1.811550240320353
```