

Division: Moments of Life (MOL)

Team: Virtual Intelligent Chat Assistant (VICA)

Team Overview

At VICA, we are developing a new platform for agencies to onboard a wave of new virtual assistants for government services. This platform will eventually replace all “Ask Jamie” chatbots and enable the agencies to onboard and collaborate effectively on it, so that they can keep up with the questions by the citizens and businesses in the long term.

Assessment Overview

Depending on the role you are signing up for, you will be assessed on part or whole of the tech stack that we are using at VICA. Please select the assessment that is relevant to the role you are applying for.

1. React frontend development
2. Node.js/Python backend development
3. Full stack development (React + Node.js/Python)
4. Data science challenge

Submission Guidelines

Start a new repository in GitHub and send the link in an email to the assessors. In your repository, make sure that you make small and frequent commits. Each commit should have a descriptive message and clean code. If you have assumptions made about the requirements of the assessment, please write that in a section in a README file.

To ensure the integrity of this assessment, do not discuss and share the work with current candidates or anyone else. Please do not host this application on AWS, GCP, Azure, DigitalOcean, Heroku or other cloud infrastructure.

Data Science Assessment

Background

You are working for an insurance company in Wakanda. The company provides health and term life insurance products. The company has compiled the data of 5,500 insurees into the file **insurance_data.csv**, to understand whether they renew their policies with the company or not. You can find the description of the data in the table below.

Task 1: You have to create a prediction model to understand insuree renewal behaviour. Create a presentation/Jupyter notebook to highlight salient features of the data and model results.

Task 2: You have to push the original data in a structured format into MongoDB. The following structure is to be followed.

Data Structure	Data Description
<pre>{ insuree#: int gender: str is45orOlder: bool isMarried: bool hasKids: bool insuredMonths: int termLifeInsurance: { hasPolicy: bool hasMultiplePolicies: bool } healthInsurance: { hasPolicy: bool riders: List[int] } premiumFrequency: int eStatements: bool monthlyPremium: float totalPremium: float renewal: bool }</pre>	<p>ID of the insuree</p> <p>M or F</p> <p>Is insuree ≥ 45 years old?</p> <p>Is insuree married?</p> <p>Does the insuree have kids?</p> <p>Months of active insurance</p> <p>Does insuree have term life policy?</p> <p>Does insuree hold multiple term life policies?</p> <p>Does insuree have health insurance?</p> <p>Does insuree have riders on this health policy?</p> <p>Premium due monthly, quarterly, annually</p> <p>Opted in for e-statements and e-policies?</p> <p>Premium amount monthly</p> <p>Total premium amount</p> <p>Does insuree renew at next premium cycle?</p>

Guiding Questions

What is the data quality and which steps did you take to clean and structure the dataset (e.g., how did you handle outliers, missing values, etc.)?

For validation purposes the company would like to see some typical cross sections of the data. For example, think of the following:

1. What is the average number of riders health insurance customers purchase, and which ones are the most popular?
2. What is the most popular rider for older people?
3. Which model did you use to gain insight into the insuree behaviour, and what are the determining factors for choosing this model?
4. Why are the insurees not renewing and what are the most important factors?

Assessment Criteria for Task 1

Data pre-processing	30%
Identifying correlations between data attributes	30%
Model selection and explanation	20%
Data visualisation	10%
Readability of Jupyter notebook	10%

Assessment Criteria for Task 2

Data modelling	30%
Environment setup	30%
Database connection and data query	30%
Code cleanliness	10%