CAPSTONE PROJECT

DEEP DIVE INTO SALVEO INNOVATIONS

PROFESSOR: RICHARD BOIRE

Group members:

-Parul Goyal (101436249)

-Shreyal Jain (101490443)

-Nischay Kiran Yedery (101457324)

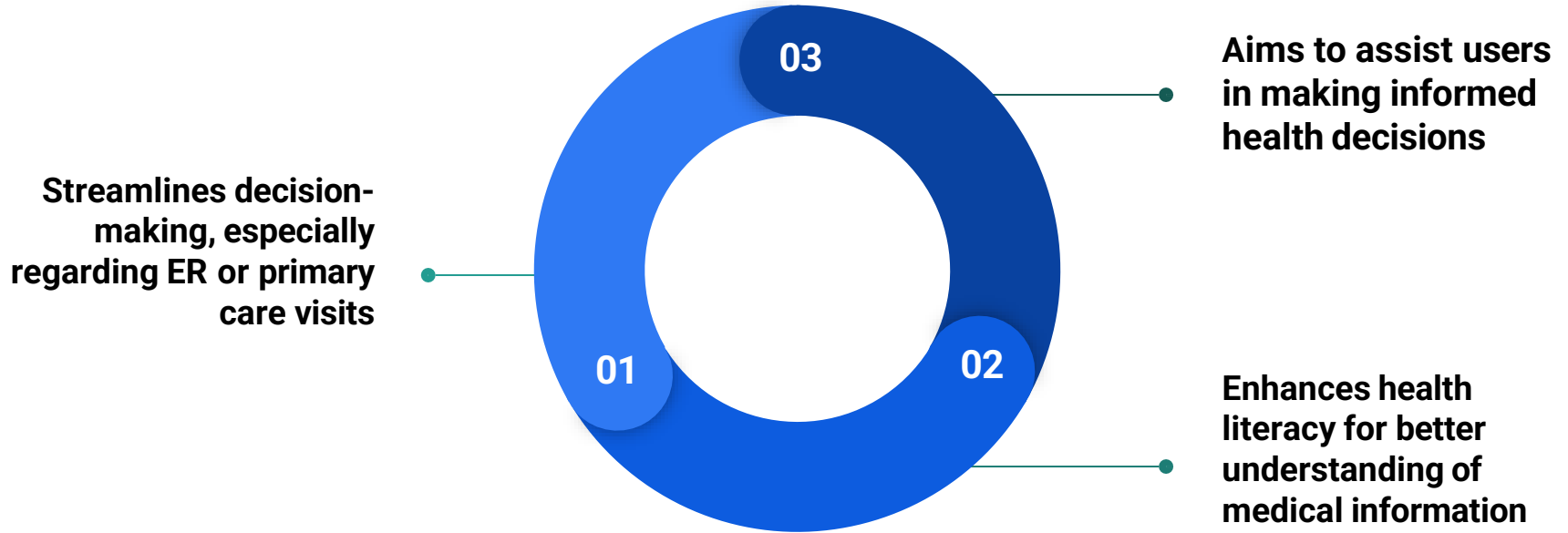**-Vishrut Shah (101453154)**

-Veron Kerr (101436413)

BACKGROUND

- US-based health tech startup, established in Washington, DC

- Develops patient-centered solutions to enhance health literacy

- 'Tyron' -a health application for easy access to vital health information

ABOUT TYRON

03 Aims to assist users in making informed health decisions

01 Streamlines decision-making, especially regarding ER or primary care visits

02 Enhances health literacy for better understanding of medical information

PROBLEM STATEMENT

To analyse and uncover gaps in understanding key demographics—age, household income, gender, employment, and education.

This knowledge deficit hampers strategic innovation for healthcare improvement.

Goal is to bridge these gaps, enabling precise, data-driven solutions aligned with Salveo's vision for transformative healthcare.
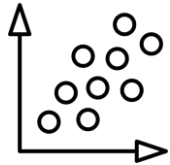
METHODOLOGY

Problem identification and objective

Identifying data and cleaning through Python

Data analysis and correlation

Gain insights and identify patterns through PowerBI

Problem solution and recommendation

DATA AUDIT

- There are about 293 records and 24 columns present in Language self.

- 255 records are present in Language caregivers. And 230 records for utility test

- The column name remains same for the rest of the 2 files as well

- Column consists of (numeric) Sample Id, Age,(Text data) Gender, Employment, Education, HHI, Device, State, Region, Industry, Race, Hispanic, Which detailed description of diabetes do you prefer?

- Each record specifies the demographics of each patients and their current condition.

OBSERVED DATASETS

Language_caregivers Dataset

- Survey filled by caregivers of individuals with hypertension
- Captures caregivers' understanding of hypertension.

Language_Self Dataset

- Responses from individuals suffering from hypertension
- Provides insights into how individuals personally understand hypertension

Utility_Test Dataset

- Collected from the landing page of Salveo Innovations
- Captures user perspectives on the beta version of the application

RESEARCH OBJECTIVE

- **To understand responses as binary (0's and 1's) into answer questions like "Are you currently providing care for individuals with chronic conditions?"**
- **Determine demographics for targeting**
- **Analyze data for suggesting improvements in specific areas.**
- **Calculated correlations and coefficients to study tangible variables**

FINDINGS FROM EACH DATASET

Language_Caregivers' Dataset

Age and Employment
- **Weak positive correlation.**
- **Slight variation in employment with age.**

Employment and Industry
- **Weak positive correlation.**
- **Marginal relation between employment type and industry.**

Household Income and Employment
- **Weak negative correlation.**
- **Inverse relationship between employment status and household income.**

Other Variables
- **Very weak or negligible correlations observed.**
- **Examples include: Region and Industry/ Device and Education**
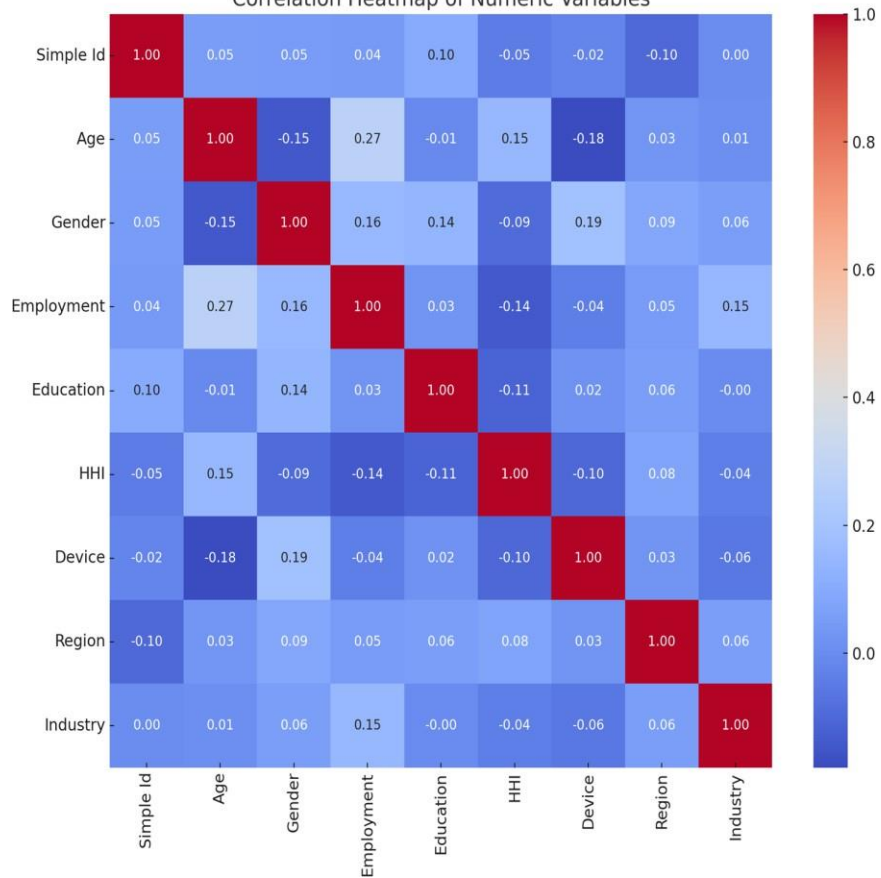
Weak positive correlation between Age and Employment (Coefficient: 0.2737).
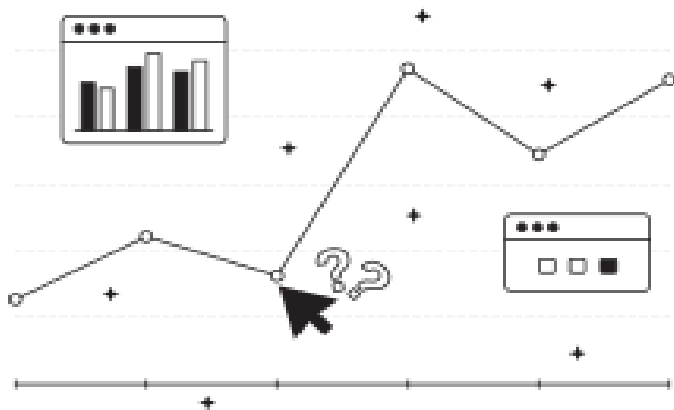
Employment and Industry also show a weak positive correlation (Coefficient: 0.1510).

Inverse relation in Household Income and Employment (Coefficient: -0.1393).

Minimal correlations in other variables like Region and Industry, Device and Education, Gender and Employment.

Utility Test Dataset

- Diverse Correlations
- Limited Linear Dependencies
- Subtle Trends

Language Self Dataset

- The dataset shows predominantly weak correlations, emphasizing the nuanced nature of the relationships between variables.
- Indicates a moderate positive correlation between age and employment, suggesting a discernible trend of improved employment with advancing age
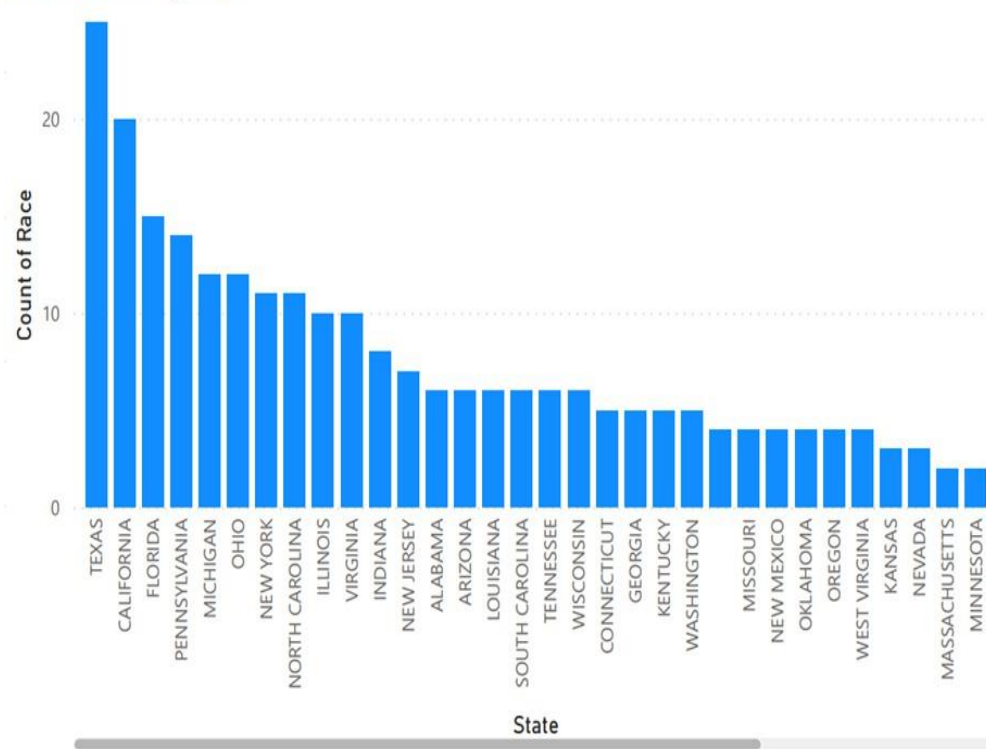
EDA REPORT

Count of HHI by Region

4 (1.54%)
40 (15.44%)
107 (41.31%)
48 (18.53%)
60 (23.17%)
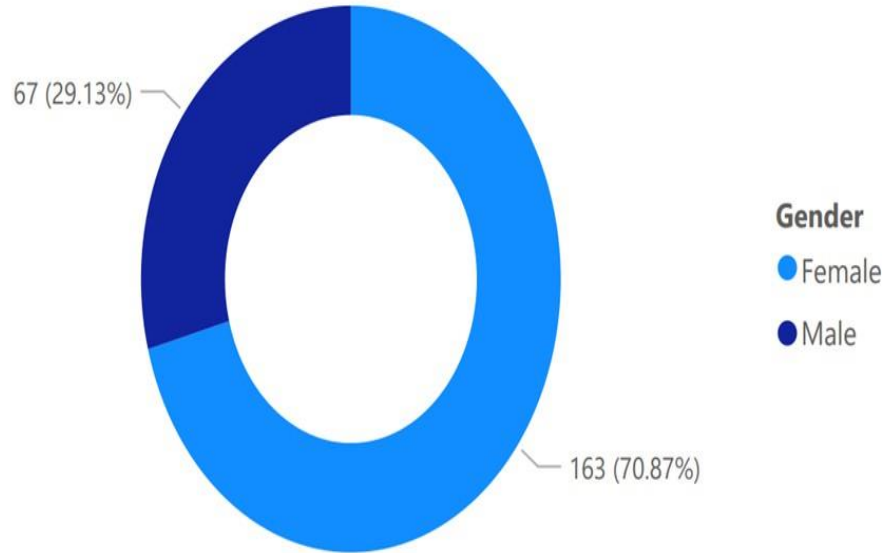
Region
- South
- Midwest
- West
- Northeast
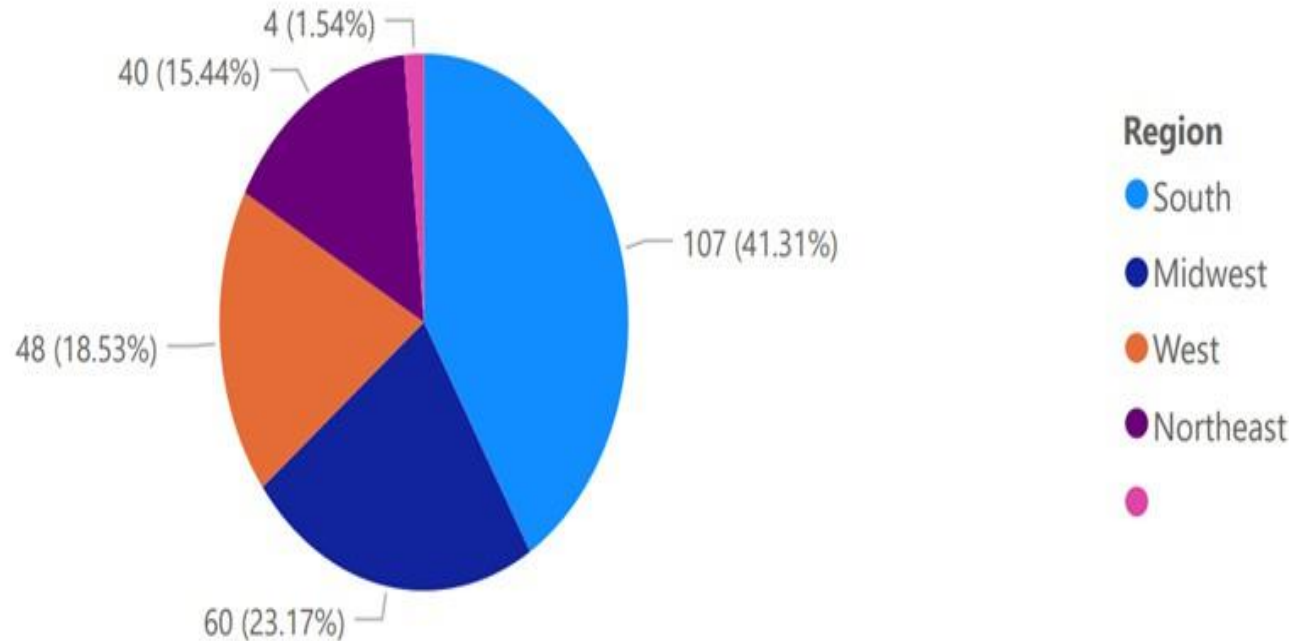
The South and Midwest boast the highest household incomes, while the West and Northeast trail behind.

Count of Race by State

Texas led in survey respondents due to its sizable population, while Minnesota had the fewest; California and Florida ranked second and third in respondent numbers
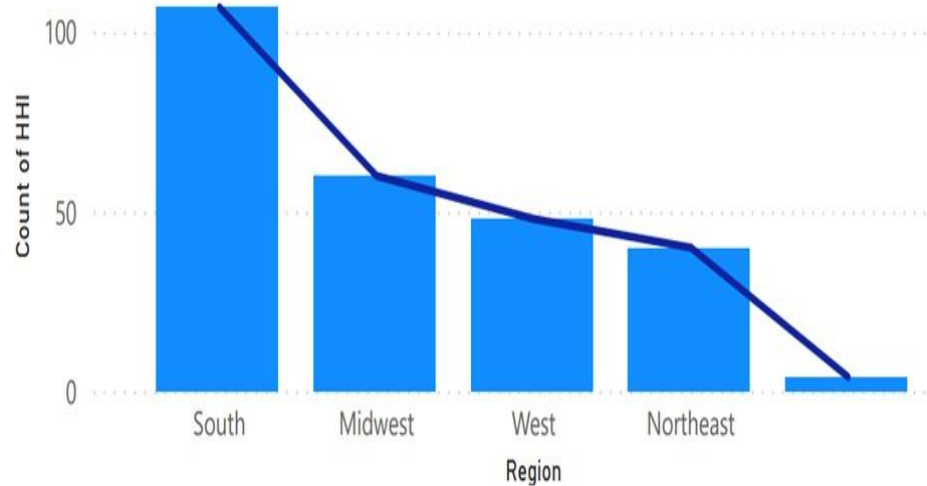
Count of Gender by Gender

67 (29.13%)

163 (70.87%)

Gender
- Female
- Male

Among the 230 survey respondents, 163 identified as female, providing valuable insights into the demographic composition, with the remaining respondents identifying as male.

Count of HHI and Count of Device by Region

● Count of HHI ● Count of Device

- Texas exhibited the highest device usage in the Southern region, aligning with its leading respondent count.

- While the Northeast and West displayed comparatively lower device counts in the survey analysis

KEY FINDINGS

CORRELATION

Age-employment and Employment-Industry have weak correlation. Inverse Relation between Household income and Employment.

EDA REPORT

The company needs to focus more on regions such as West and NorthEast so that more people are aware of the application.

ANALYSIS

People in states such as Minnesota, Nevada and Massachusetts need to be made aware of the application by advertising the application through billboards or through radios.

KEY CHALLENGES

1. Disease Definition Comprehension:
- Client challenge: Understanding the disease definition.
 - Difficulty in grasping the nature of the disease they are tackling.

2. Household Income Entry Difficulty:
 - Challenge: Entering precise Household Income data.
 - Difficulty in providing specific data for better situational analysis.

3. Sentiment Analysis for Textual Data:
 - Focus on:
- Analyzing text data.
- Understanding user sentiments regarding various types of questions.

PROCESS

1. Numeric Conversion for Definition Analysis

2. Income Categorization for Medical Precision

3. Sentiment Analysis for User Understanding
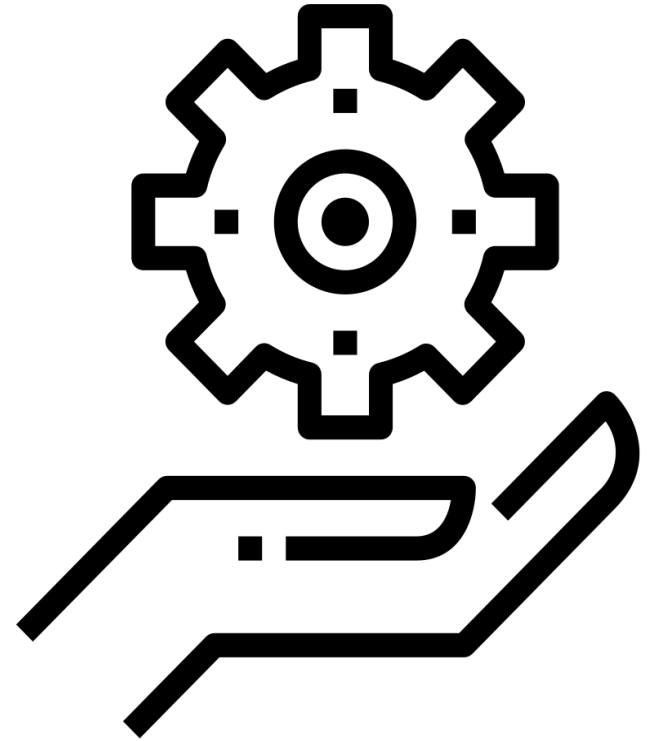
SOLUTIONS

Enhanced Data Entry Interface:
-Implement drop-down options for data entry.
-Facilitates easy selection based on user understanding.
-Encourage responses written for universal comprehension.

Numeric Labeling for Efficient Analysis:
-Assign numbers to data entries in next to text.
-Simplifies data analysis.
-Accelerates insight generation for company growth.