Kau William

Langelier Maxime

Leorat Marius

Lefrancois Gabin

Klink Carl

Girardot Arthur

**Projet Life Insurance**

Explainability Ai

**Table of Contents**

Table of Contents

[1. Introduction 3](#_Toc193563223)

[1.1 Project Context 3](#_Toc193563224)

[1.2 Objectives 3](#_Toc193563225)

[2. Data Collection & Integration 4](#_Toc193563226)

[2.1 Data Sources 4](#_Toc193563227)

[2.2 Data Preparation & Cleaning 4](#_Toc193563228)

[3. Data Analysis & Insights 5](#_Toc193563229)

[3.1 Key metrics & Trends 5](#_Toc193563230)

[3.2 Data Modeling 6](#_Toc193563231)

[4. Dashboard Design & Functionalities 6](#_Toc193563232)

[4.1 Visualization Requirements 6](#_Toc193563233)

[4.2 User Experience Considerations 6](#_Toc193563234)

[5. Technical Requirements 7](#_Toc193563235)

[5.1 Tools & Technologies 7](#_Toc193563236)

[5.2 Security & Compliance 7](#_Toc193563237)

[6. Project Timeline & Milestones 7](#_Toc193563238)

[6.1 Key Deliverables & Deadlines 7](#_Toc193563239)

[7. Conclusion 8](#_Toc193563240)

# 1. Introduction

## 1.1 Project Context

LifeSure Insurance is a prominent player in the insurance industry, facing increasing pressure from both clients and regulatory bodies to incorporate sustainability and social responsibility into its business model. As the market evolves, customer expectations shift towards modern, ethical, and sustainable insurance policies. In response, LifeSure Insurance has decided to transition from traditional policies to innovative, customer-centric solutions that align with these contemporary demands.

To achieve this transformation, LifeSure Insurance requires an advanced data visualization tool that provides actionable insights into customer behaviors, preferences, and expectations. By leveraging data-driven decision-making, the company aims to develop sustainable policies that resonate with its target audience, ensuring long-term growth and competitiveness in the industry.

## 1.2 Objectives

The primary objective of this project is to design and develop an advanced data visualization tool that enables LifeSure Insurance to:

* Analyze customer behaviors, needs, and expectations to guide policy development.
* Identify emerging trends in sustainability and social responsibility within the insurance sector.
* Enhance decision-making by providing intuitive and interactive visual representations of data.
* Support the company's transition towards modern, customer-focused, and sustainable insurance policies.
* Improve customer engagement and satisfaction through data-driven insights.

This project aims to provide a comprehensive, data-driven approach to help LifeSure Insurance navigate the evolving landscape of the insurance industry while fostering sustainability and customer satisfaction.

# 2. Data Collection & Integration

## 2.1 Data Sources

To develop an effective data visualization tool, this project will utilize multiple datasets that provide insights into customer segmentation, insurance claims, and corporate environmental impact. The datasets used include:

* [Insurance Claim Analysis](https://www.kaggle.com/datasets/thedevastator/insurance-claim-analysis-demographic-and-health): Contains demographic and health-related insurance claim data, which helps analyze customer claims patterns.
* [Customer Segmentation Data](https://www.kaggle.com/datasets/ravalsmit/customer-segmentation-data): Provides information on customer demographics, behaviors, and purchasing patterns.
* [Corporate Environmental Impact](https://www.kaggle.com/datasets/mannmann2/corporate-environmental-impact): Offers data on corporate sustainability efforts and environmental impact.
* [Caravan Insurance Challenge](https://www.kaggle.com/datasets/uciml/caravan-insurance-challenge): A dataset for predicting insurance policy purchases based on customer attributes.

By integrating these datasets, the project aims to uncover valuable insights that will enable LifeSure Insurance to create data-driven, sustainable, and customer-focused insurance policies.

## 2.2 Data Preparation & Cleaning

Before utilizing the datasets, a comprehensive data cleaning and preparation process will be undertaken to ensure accuracy and reliability. This includes:

* Checking for and handling missing or null values to prevent inconsistencies in the analysis.
* Removing extreme values or outliers that could skew the insights and mislead decision-making.
* Deleting unused columns that do not contribute to the analysis to optimize data processing efficiency.
* Standardizing data formats and ensuring consistency across datasets for seamless integration.
* Handling duplicate entries to maintain data integrity.
* Performing exploratory data analysis (EDA) to identify potential patterns and anomalies before visualization.

These steps will ensure that the data used in the visualization tool is clean, structured, and ready for meaningful analysis, ultimately supporting LifeSure Insurance in making informed, strategic decisions.

# 3. Data Analysis & Insights

## 3.1 Key metrics & Trends

 **Customer Demographics & Segmentation**

* **Age Distribution & Headcount Analysis**  
  • Detailed statistical summaries (mean, median, variance) from Dataset 1 and Dataset 3 reveal distinct age clusters.  
  • Visualizations highlight peaks in younger (mid 20s and mid 30s) and older (mid 50s and 60+) demographics.  
  • These trends help pinpoint the age groups with the highest insurance uptake and claim frequency.
* **Nationality Distribution**  
  • Analysis of customer nationalities identifies the top 10 most frequent groups.  
  • Frequency counts and percentage shares provide insight into geographic and cultural market segments.  
  • This segmentation is crucial for targeted marketing and risk profiling.
* **Customer Segmentation Insights**  
  • Merged data from multiple sources allowed the creation of distinct customer segments.  
  • The segmentation strategy is based on demographic attributes and behavioral patterns extracted from the datasets.  
  • Visual breakdowns (bar graph) illustrate how segments differ in size and risk profile.

 **Insurance Claims & Health Trends**

* **Claim Amount Distribution**  
  • Detailed histograms and box plots display the distribution of claim amounts, emphasizing the range and outliers.  
  • Key statistical measures (mean claim amount, standard deviation) indicate overall claim variability, with a lot more claims ranging from 0 to 15000.  
  • This analysis informs premium pricing and risk management strategies.
* **Smoking vs. Non-Smoking Analysis**  
  • Comparative analysis between patients who smoke and those who do not shows clear differences in claim frequency and average claim size.  
  • Correlation tests and regression analyses from the notebook underscore the impact of smoking on claim risk.  
  • These insights support the refinement of underwriting criteria.

 **Environmental Impact Metrics**

* **Environmental Intensity Distribution**  
  • Analysis of corporate environmental impact data reveals the spread and central tendencies of environmental intensity metrics.  
  • Trends indicate how environmental factors correlate with customer behavior or risk profiles in the insurance context.  
  • This metric is integrated to support corporate social responsibility (CSR) evaluations and sustainable business practices.

 **Data Processing & Integration Methodology**

* **Data Cleaning and Merging**  
  • Extensive data cleaning steps were applied to each dataset (e.g., handling missing values, outlier removal, standardization of fields).  
  • Merging multiple datasets from diverse sources (insurance claims, customer segmentation, environmental impact, and the Caravan Insurance Challenge) ensured a richer, more robust dataset.  
  • The integration allowed cross-validation of metrics, thereby enhancing the reliability of the trends extracted.

 **Statistical and Visual Insights**

* **Descriptive Statistics and Visual Analytics**  
  • The notebook’s code includes generating detailed statistical summaries for each dataset, forming the basis for the visual trends presented.  
  • Advanced visualizations (e.g., scatter plots, line graphs) were used to track correlations between demographics and claim amounts.  
  • These plots serve as key tools in identifying anomalies and trends over different customer segments.

## 3.2 Data Modeling

# 4. Dashboard Design & Functionalities

## 4.1 Visualization Requirements

• The dashboard includes visualizations created using Python’s plotting libraries.  
• Visualizations consist of histplots, countplots, and barplots (with some plots enhanced using kde curves), offering a clear view of distributions and categorical frequencies.  
• These charts support the presentation of key metrics such as customer age distributions, claim amounts, and nationality frequencies.

## 4.2 User Experience Considerations

• The dashboard is designed for the fictional insurance company LifeSure, targeting decision makers within the organization.  
• Navigation and layout principles have been kept simple and direct to ensure ease of understanding, though further enhancements (e.g., responsive design or accessibility features) are open for additional input from the team.

# 5. Technical Requirements

## 5.1 Tools & Technologies

## 5.2 Security & Compliance

# 6. Project Timeline & Milestones

The project is divided into three main phases, with specific deadlines for intermediary and final deliverables.

**Phase 1: Data Collection and Integration *(Early Development Stage)***

* Identify and collect relevant external datasets (e.g., customer demographics, environmental data, market trends).
* Perform data cleaning, preprocessing, and merging to create a unified dataset.
* Ensure data standardization and integrity for further analysis.

**Phase 2: Data Analysis and Specification Development *(Midway Progress)***

* Conduct Exploratory Data Analysis (EDA) to uncover key trends, patterns, and customer segments.
* Develop statistical models and identify correlations between variables (e.g., sustainability concerns, customer preferences, claim behavior).
* Document the key findings and define specifications for the dashboard.

**Phase 3: Visualization and Communication of Results *(Final Development & Presentation Stage)***

* Design and develop an interactive executive-level dashboard based on analyzed insights.
* Finalize the visual presentation of key metrics for decision-makers.
* Prepare final documentation and business presentation summarizing findings, methodology, and strategic recommendations.

## 6.1 Key Deliverables & Deadlines

* **Intermediary Deliverable:** *(Checkpoint Submission)*
  + Due **March 6th, 10:00 AM**
  + Purpose: Provide a progress update, allowing feedback from the jury to refine the final submission.
* **Final Deliverable:** *(Complete Submission)*
  + Due **March 25th, 10:00 AM**
  + Includes:
    - Final dashboard and its source files
    - Complete project documentation
    - Specification document detailing methodology and findings
* **Oral Presentation:** *(Evaluation & Defense)*
  + **Duration:** 45 minutes per group (30 minutes presentation + 15 minutes Q&A).
  + **Submission of Presentation Slides:** Must be sent no later than **one hour after the oral presentation.**

.

# 7. Conclusion