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 2](#_Toc724217814)

[1.1 Project Context 2](#_Toc1117473703)

[1.2 Objectives 2](#_Toc386916427)

[2. Data Collection & Integration 3](#_Toc1301905080)

[2.1 Data Sources 3](#_Toc1753409470)

[2.2 Data Preparation & Cleaning 3](#_Toc1661312324)

[3. Data Analysis & Insights 4](#_Toc1337519716)

[3.1 Key metrics & Trends 4](#_Toc498473791)

[4. Dashboard Design & Functionalities 5](#_Toc842296357)

[4.1 Visualization Requirements 5](#_Toc318512606)

[4.2 User Experience Considerations 5](#_Toc1471213165)

[5. Project Timeline & Milestones 5](#_Toc11171816)

[5.1 Key Deliverables & Deadlines 5](#_Toc2080813220)

[6. Conclusion 5](#_Toc1122781264)

# 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.
* [Disaster Dataset](https://www.kaggle.com/datasets/brsdincer/all-natural-disasters-19002021-eosdis): Contains historical records of natural disasters, revealing trends in frequency and severity that inform risk assessment and support innovative policy development.

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**  
  The analysis of customer demographics highlights key trends in age distribution, nationality, and segmentation. Statistical summaries from Dataset 1 and Dataset 3 reveal distinct age clusters, with peaks in the mid-20s, mid-30s, mid-50s, and 60+. These insights help identify the age groups with the highest insurance uptake and claim frequency.
* **Nationality Distribution**  
  Nationality distribution analysis focuses on the top 10 most common customer groups. Frequency counts and percentage shares provide a clearer understanding of geographic and cultural market segments, which are essential for targeted marketing and risk profiling.
* **Customer Segmentation Insights**  
  By merging multiple data sources, distinct customer segments were created based on demographic attributes and behavioral patterns. Visual breakdowns, such as bar graphs, illustrate differences in segment size and risk profiles, offering valuable insights for strategic decision-making.

 **Insurance Claims & Health Trends**

* **Claim Amount Distribution**  
  The analysis of claim amount distribution provides key insights into variability and outliers. Histograms and box plots highlight the range of claim amounts, with a significant concentration of claims between 0 and 15,000. Key statistical measures, such as the mean claim amount and standard deviation, help assess overall claim variability, aiding in premium pricing and risk management strategies.
* **Smoking vs. Non-Smoking Analysis**  
  A comparative analysis of smoking and non-smoking patients reveals notable differences in claim frequency and average claim size. Correlation tests and regression analyses demonstrate the impact of smoking on claim risk. These findings play a crucial role in refining underwriting criteria to better assess risk and adjust policy terms accordingly.

· **Environmental Impact Metrics**

* **Environmental Intensity Distribution**  
   The analysis of corporate environmental impact data highlights the distribution and central tendencies of environmental intensity metrics. Trends reveal how environmental factors correlate with customer behavior and risk profiles within the insurance context.
* The analysis also includes an environmental score, allowing for direct comparisons between companies and countries. This score helps identify the businesses and regions with the highest environmental impact, providing valuable insights for sustainability assessments. By ranking companies based on their environmental intensity, we can pinpoint key areas for improvement and align insurance strategies with sustainability objectives.
* Environmental impact metrics are enhanced by the disaster dataset analysis, which reveals that natural disasters have grown both more frequent and deadlier from 1970 to 2021. This increase, driven largely by climate change and environmental degradation, highlights the growing risks for communities—especially homeowners in high-risk areas, young families, and older individuals. These trends suggest that LifeSure should take disaster data into account when evaluating environmental impacts, ensuring that its insurance policies are robust and aligned with sustainable practices. As public awareness of environmental issues grows, integrating these metrics will help guide the development of effective policies.

# 4. Dashboard Design & Functionalities

## 4.1 Visualization Requirements

Initially, data visualizations were created directly in the notebook using Python’s plotting libraries to explore trends and identify the most relevant graphs for analysis. This step allowed for a quick evaluation of different visualization techniques, helping to determine which charts provided the most meaningful insights. By testing various approaches, such as histograms, count plots, and bar plots, the analysis was refined to highlight key trends effectively.

## 4.2 User Experience Considerations

Once the most insightful visualizations were identified, the process shifted to Power BI to create interactive dashboards. This transition enabled a more dynamic and user-friendly presentation of key metrics, including customer age distributions, claim amounts, and nationality frequencies. Some plots were enhanced with KDE curves to provide smoother distributions and deeper insights. The Power BI dashboard improved data accessibility, allowing stakeholders to explore trends more intuitively and make informed decisions based on real-time insights.

## 4.3 Functionalities

Using Power BI, we created three dashboards, one for each dataset. In df1, we used bar plots to look at the most common income segment of clients. We also found out that the most common age group for clients were the 30s, which led us to analyze the income level by education level for clients aged between 30 and 39 years old using a box plot. For df3, we used a heatmap to look at the countries with the highest occurrence of natural disasters. We then looked at the total population affected by these disasters per continent, and also the disaster count by type in Europe from 1970 to 2021. In df4, we used a correlation matrix between the parameters Age, BMI, Children, Claim and Risk Score, using python scripts. We also used scatterplots for Risk Score by BMI and number of claims by BMI. We also tried to look at a correlation between smoking and high BMI using bar plots.

# 5. Conclusion

This project successfully developed an advanced data visualization tool to support LifeSure Insurance in transitioning toward customer-centric, sustainable policies. By integrating multiple datasets, including insurance claims, customer segmentation, corporate environmental impact, and disaster data, we identified key insights that enhance decision-making and risk management.

Our analysis revealed critical customer trends, such as dominant age segments and nationality distributions, helping LifeSure tailor its policies to different demographic groups. Insights from claim distributions and health factors, such as the impact of smoking and BMI on claim frequency, provide valuable guidance for refining underwriting criteria. Additionally, the study of environmental impact metrics highlighted the increasing risk posed by climate change and natural disasters, reinforcing the importance of sustainability in insurance offerings.

The interactive Power BI dashboards enhanced data accessibility, allowing stakeholders to explore key insights dynamically. By leveraging visualizations such as histograms, heatmaps, and correlation matrices, the tool provides an intuitive way to assess customer behaviors, policy risks, and sustainability trends.

In conclusion, this project equips LifeSure Insurance with the data-driven capabilities needed to adapt to evolving market demands. By utilizing advanced analytics and visualization techniques, the company can develop innovative insurance solutions that align with customer expectations, regulatory requirements, and sustainability objectives, ensuring long-term growth and competitiveness in the industry.