**Contents**

[a) Purpose 1](#_Toc175085751)

[b) Intended Audience and Use 1](#_Toc175085752)

[c) Product Scope 3](#_Toc175085753)

[d) Definitions and Acronyms 4](#_Toc175085754)

[Overall Description 4](#_Toc175085755)

[User Needs 4](#_Toc175085756)

[Assumptions and Dependencies 5](#_Toc175085757)

[System Features and Requirements 5](#_Toc175085758)

[Functional Requirements 5](#_Toc175085759)

[External Interface Requirements 5](#_Toc175085760)

[ User 5](#_Toc175085761)

[ Hardware 5](#_Toc175085762)

[ Software 5](#_Toc175085763)

[ Communications 5](#_Toc175085764)

[System Features 5](#_Toc175085765)

[Nonfunctional Requirements 5](#_Toc175085766)

[ Performance requirements 5](#_Toc175085767)

[ Safety requirements 5](#_Toc175085768)

[ Security requirements 5](#_Toc175085769)

[ Usability requirements 5](#_Toc175085770)

[ Scalability requirements 5](#_Toc175085771)

# Introduction:

In the highly competitive health care insurance industry, understanding customer behavior and optimizing revenue streams are critical to maintaining a competitive edge. The integration of Big Data analytics provides an opportunity to gain valuable insights into customer behavior, competitor strategies, and market trends. By analyzing data from various sources, including competitor data, the company can develop customized insurance offers, optimize existing policies, and accurately calculate royalties for past customers. This project aims to enhance the company’s revenue and customer engagement by leveraging the power of Big Data.

## Purpose

The primary purpose of this project is to design and implement data pipelines that will enable the company to develop effective business strategies. These strategies will be informed by in-depth analysis of customer behavior, optimization of insurance offers, and accurate calculation of royalties. The project seeks to harness the potential of Big Data analytics to:

* Track customer behavior and conditions to tailor insurance policies.
* Analyze competitor data to identify trends and opportunities.
* Calculate royalties for past customers based on historical data.
* Ultimately, enhance revenue through data-driven decision-making.

## Intended Audience and Use

This document is intended for a diverse audience, including:

* **Stakeholders:** Executives, project sponsors, and decision-makers who need a clear understanding of the project's objectives, scope, and expected outcomes.
* **Data Analysts:** Professionals responsible for designing, implementing, and maintaining the data pipelines and analytics processes.
* **Business Strategists:** Individuals tasked with interpreting the data insights to formulate and execute business strategies aimed at revenue enhancement and customer satisfaction.

**How the Audience Will Use This Document:**

* **Stakeholders** will use the document to align the project goals with the company’s broader strategic objectives and to monitor progress.
* **Data Analysts** will reference the document to ensure their work aligns with the project’s goals and scope, particularly in developing data models and algorithms.
* **Business Strategists** will leverage the insights provided to shape marketing campaigns, policy offers, and other revenue-generating initiatives.

## Product Scope

The scope of this project encompasses several key areas critical to achieving the desired outcomes. These areas include:

1. **Identifying the Disease with the Maximum Number of Claims:**
   * Analyze claims data to determine which diseases are most frequently claimed.
   * Use insights to adjust policy offerings and pricing strategies.
2. **Analyzing Subscriber Demographics:**
   * Segment the subscriber base by age, location, income, and other relevant demographics.
   * Understand the needs and behaviors of different demographic groups to tailor insurance products accordingly.
3. **Evaluating the Profitability of Different Policy Groups:**
   * Assess the financial performance of various insurance policy groups.
   * Identify underperforming policies and recommend adjustments to improve profitability.
4. **Tracking Claims Based on Various Criteria:**
   * Monitor claims data based on factors such as time, location, and disease type.
   * Use this information to detect patterns, predict future claims, and optimize risk management strategies.

## Definitions and Acronyms

This section provides definitions and explanations for key terms and acronyms used throughout the document:

* **Big Data Ecosystem:** The complex infrastructure of data collection, storage, processing, and analysis that supports large-scale data analytics.
* **ROI (Return on Investment):** A performance measure used to evaluate the efficiency of an investment or compare the efficiency of several investments.
* **Subscriber:** An individual who holds an insurance policy with the company.
* **Subgroup:** A specific segment of the subscriber base, categorized by factors such as demographics or policy type.
* **Claim:** A request made by the policyholder for payment of the benefits provided by the insurance policy.
* **Data Pipeline:** A series of data processing steps, often involving extraction, transformation, and loading (ETL) of data from various sources into a central repository for analysis.

# Overall Description

## User Needs

## Assumptions and Dependencies

# System Features and Requirements

## Functional Requirements

## External Interface Requirements

### User

### Hardware

### Software

### Communications

## System Features

## Nonfunctional Requirements

### Performance requirements

### Safety requirements

### Security requirements

### Usability requirements

### Scalability requirements