How do Insurance policy sales, charges, and claim patterns vary across regions, age groups, and genders, and what insights can help optimize product offerings and claims management?

CAPSTONE PROJECT

FINAL REPORT

Valluri Venkatram

CONTENTS

1. Brief Introduction About the Problem Statement and Need to Solve It4
a) Overview of the insurance landscape and key challenges
b) Introduction to problem statement
c) Problem understanding
2. Importing the Data from Source
a) Identifying data sources (e.g., databases, CSV files)
b) Methods for data extraction (e.g., ETL processes)
c) Tools and technologies used for data import (e.g., Power BI, SQL)
3. Data Cleaning and Preprocessing7
a) Handling missing values (e.g., imputation or removal)
b) Standardizing data formats (e.g., date formats, text casing)
c) Removing duplicates and inconsistencies
d) Normalizing data (e.g., scaling numerical values)
e) Creating custom columns
4. Creating Relationships Between Tables9
a) Identifying primary and foreign keys in tables
b) Establishing relationships (e.g., One to many, many to many)
c) Validating relationships to ensure data integrity
5. Build Calculated Measures Using DAX10
a) Defining key metrics and calculations
b) Writing DAX formulas for calculated measures
c) Creating measures for specific analyses
6. Create Different Types of Charts, Tables and Use Slicers and Filters
Effectively
a) Selecting appropriate chart types (e.g., bar charts, pie charts, line graphs)
b) Implementing slicers for interactive filtering (e.g., by age, gender, region)
c) Using filters to refine data views and enhance user experience

	sign Interactive Dashboards23 Layout and organization of dashboard elements
	Incorporating design principles for user friendly navigation Adding interactivity Customizing visuals for branding and aesthetics
c)	
	ions
	Interpreting dashboard visualizations and data patterns
,	Identifying key trends and anomalies
	Generating reports summarizing findings Making recommendations based on insights for strategic planning
ŕ	

LIST OF FIGURES

Plan Wise Insurance Analysis Dashboard

- 1. Donut Chart Total Insurance Policy by Plan Name
- 2. Bar Chart Total Charges by Plan Name
- 3. Line Chart Total Insurance Policy by Plan Name and Gender Status
- 4. Stacked Bar Chart Total Insurance Policy by Plan Name and Age Group
- 5. Line Chart Claim Policy and Nonclaim Policy by Plan Name

Region Wise Insurance Analysis Dashboard

- 1. Donut Chart Total Insurance Policy by Region
- 2. Bar Chart Total Charges by Region
- 3. Line Chart Total Insurance Policy by Region and Gender Status
- 4. Stacked Bar Chart Total Insurance Policy by Region and Age Group
- 5. Line Chart Claim Policy and Nonclaim Policy by Region

Combined Dashboard (Plan + Region Filters)

- 1. Donut Chart Total Insurance Policy by Region
- 2. Bar Chart Total Charges by Region
- 3. Line Chart Gender based Insurance Policy Distribution by Region
- 4. Stacked Bar Chart Age Group Distribution of Insurance Policies by Region
- 5. Line Chart Comparative Analysis of Claim vs. Nonclaim Policies by Region

Executive Summary

This set of dashboards provides a comprehensive analysis of insurance policies across different plans and regions. It offers insights into policy distribution based on gender, age group, and claim status, alongside total charges incurred by various segments. With interactive filters, users can explore specific plans, regions, or policy IDs. This visualization aids stakeholders in identifying trends and optimizing offerings. The dashboards empower decisionmaking by revealing key performance indicators across plans and regions.

Introduction to Problem Statement

Insurance companies face the challenge of managing diverse policies across multiple plans and regions while addressing customer needs efficiently. With numerous factors like age group, gender, and claim patterns influencing policy uptake, it becomes essential to track performance across various dimensions. Additionally, understanding the charges generated and identifying areas of improvement in claim management is critical for maximizing profitability. Insurance providers need detailed insights to align their offerings with market demands and ensure better resource allocation. The ability to analyze claims and nonclaims effectively is essential to maintaining customer satisfaction and controlling operational costs. These dashboards provide actionable insights by organizing insurance data into intuitive visualizations, helping stakeholders monitor key metrics with precision.

Problem Understanding

Problem Statement

The insurance industry requires an analytical tool that provides detailed insights into the performance of various policies segmented by plans and regions. It is necessary to visualize data such as policy distribution, charges, age group demographics, and claim patterns to facilitate better decisionmaking. Without such a dashboard, it becomes difficult to track trends, identify bottlenecks, and optimize offerings.

Need for Problem Statement

A structured analysis through interactive dashboards enables insurance providers to quickly assess how well different plans perform across demographic groups and regions. It allows them to track claim behavior and policy uptake trends, which is essential for refining strategies and enhancing operational efficiency. A lack of such insights can result in poor resource management, customer dissatisfaction, and missed growth opportunities. This problem statement highlights the critical need for advanced data visualization tools to support decisionmaking, drive profitability, and improve service delivery.

Understanding the Problem Statement

In the competitive insurance industry, it is essential to track the performance of policies across various plans, regions, and customer demographics to remain competitive and profitable. Insurance companies deal with multiple variables such as gender, age group, plan types, regionbased preferences, and claim patterns that directly influence policy sales and customer behavior. Without an organized view of these data points, companies risk losing insights into customer trends and operational gaps.

The problem revolves around the need to understand which plans and regions generate the most revenue, which demographic groups show the highest policy uptake, and how claims versus nonclaims impact profitability. Identifying such trends requires a centralized system that provides a complete view of policy performance across multiple categories, enabling better product alignment and strategic decision making. By visualizing these critical factors with dashboards, stakeholders can proactively respond to shifts in customer behavior, address inefficiencies in claims management, and enhance policy offerings to improve overall performance.

Objective:

The primary objective of the dashboard project is to enhance decision-making capabilities within the insurance sector by providing a comprehensive analysis of policy data. This includes identifying trends and patterns that impact customer behavior and policy performance. By effectively importing and preprocessing data from diverse sources, we aim to ensure data integrity and quality. Establishing meaningful relationships between tables will facilitate a more holistic view of the data. Additionally, the development of calculated measures using DAX will enable precise analytics that drive insights. Through the creation of interactive visualizations and dashboards, the goal is to empower stakeholders to explore data intuitively and derive actionable insights. Ultimately, the project seeks to support strategic initiatives and improve operational efficiency by translating complex data into clear, meaningful narratives.

Here are the tables and their columns:

Table: plan data insurance plan_id insurance_plan_name

Table: POLICY

age

AGE GROUP

bmi

charges

children

Column1

Gender Status

insrance plan_id

insurance plan_id1

insurance policy_id

insuranceclaim

region Id

sex

smoker

Table: region

Region

region Id

To fulfill the above given objectives, we need to create analysis report in Power BI desktop using the following steps:

- Import data from various sources.
- Use Power Query for data cleaning and transformation.
- Create relationships between tables.
- Filter and slice your data and use drilldown capabilities for deeper analysis.
- Build calculated columns and measures using DAX.
- Create different types of charts, tables and Use slicers and filters effectively.
- Design interactive dashboards.
- Analyze the data to identify meaningful insights and make data driven decisions.

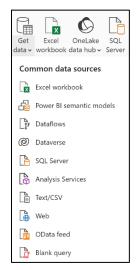
1. Import Data from Source

Objective: To bring all necessary data into Power BI for analysis.

Steps:



- Open Power BI Desktop: Launch the Power BI Desktop application.
- Get Data: Click on the "Get Data" button located on the Home ribbon.
- Choose Data Source: Select the type of data source you want to connect to (e.g., Excel, CSV, SQL Server, SharePoint, etc.). Power BI supports a wide range of data sources including cloudbased services like Azure and webbased data.



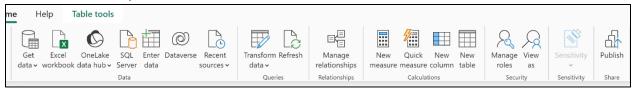
- *Connect to Data Source:* Follow the prompts to establish a connection.
- For instance, if you are importing data from an Excel file:
- Click on "Excel".
- Browse and select your Excel file.
- Click "Open".
- Load Data: In the Navigator window, select the tables or sheets you want to import and click "Load" to bring them into Power BI.
- Ensure your data is clean and wellstructured in the source files.
- Power BI can handle large datasets efficiently, but consider the size and complexity of the data you're importing.

2. Use Power Query for Data Cleaning and Transformation

Objective: To prepare and clean the data for analysis.

Steps:

• Launch Power Query Editor: After importing your data, click on "Transform Data" to open the Power Query Editor.



Data Preprocessing Overview

To ensure the integrity and usability of the datasets used for analysis, a comprehensive data preprocessing step was conducted on the following tables: plan data, POLICY, and region. The preprocessing involved the following key actions:

1. Data Cleaning:

- Removed any duplicate entries and unnecessary columns, such as `Column1` in the POLICY table, to streamline the dataset.
- Addressed missing values in critical fields like age, charges, and smoker status to prevent skewed analysis.

2. Data Transformation:

- Standardized column names across tables for consistency and clarity.
- Categorized the `age` field into defined `AGE GROUP` categories for better segmentation and analysis.
- Converted relevant columns to appropriate data types (e.g., numerical, categorical) to facilitate efficient data operations.

3. Data Integration:

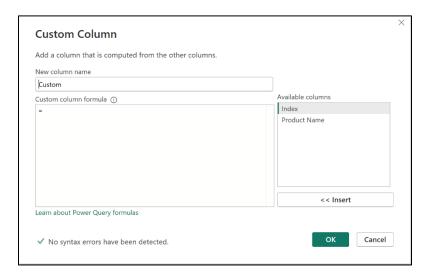
• Merged tables based on common identifiers (e.g., `insurance plan_id` and `region Id`) to create a unified dataset that supports comprehensive analysis.

4. Data Validation:

• Conducted validation checks to ensure that the data accurately reflects the source information and adheres to expected ranges and formats.

These preprocessing steps laid the foundation for accurate insights and analyses in subsequent stages of the project.

To create custom columns: Add required columns using Custom column of Add column

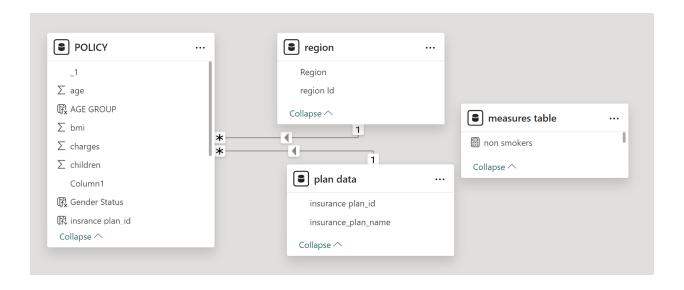


3. Create Relationships Between Tables

- ➤ Objective: To establish logical connections between different tables to enable comprehensive analysis.
- > Steps:
- Go to Model View: Click on the "Model" icon on the left sidebar to view all imported tables.



- Create Relationships:
- *Drag and Drop:* Click on a field in one table and drag it to the corresponding field in another table to create a relationship.
- *Define Relationship*: The "Manage Relationships" dialog allows you to set up relationships manually. Here, specify the primary and foreign keys.
- ➤ Cardinality and CrossFiltering:
- *Cardinality:* Define the type of relationship (onetomany, manytoone, manytomany).
- *CrossFiltering*: Set the direction of data filtering (single or both directions).
- Ensure that related fields have matching data types.



4. Build Calculated Columns and Measures Using DAX

Objective: To perform advanced calculations and derive new insights from your data.

• DAX (Data Analysis Expressions) is a powerful language for creating complex calculations and aggregations.

Steps:

- Open Data View: Click on the "Data" icon on the left sidebar to view your tables.
- Create Calculated Column:
- New Column: Click on "New Column" in the "Home" ribbon of table view.



- DAX Formula: Enter a DAX formula to define the new column.
 For example, to calculate profit margin: Profit Margin = DIVIDE([Profit], [Sales]).
- Create Measures:
- New Measure: Click on "New Measure" in the "Home" ribbon of table view.
- DAX Formula: Define a measure using DAX.
 For example, to calculate total sales: Total Sales = SUM(Sales[Amount]).
- Calculated columns are evaluated row by row, whereas measures are aggregated calculations.

Measures Created:

1. Nonsmokers

• Formula:

• `Non Smokers = CALCULATE([total insurance policy], POLICY[smoker] = 0)`

• Explanation:

• This measure calculates the total number of insurance policies where the `smoker` field is 0, indicating nonsmokers.

• Use Case:

This measure helps assess how many policyholders are nonsmokers, which could provide
insights into health risks and premium variations. It is often used to compare healthrelated
trends between smokers and nonsmokers.

• Impact:

• It helps in understanding the proportion of policies issued to nonsmokers, which could influence insurance product pricing strategies.

2. Smokers

• Formula:

• `Smokers = CALCULATE([total insurance policy], POLICY[smoker] = 1)`

• Explanation:

• This measure calculates the number of insurance policies for people identified as smokers (where `smoker = 1`).

• Use Case:

• It helps in analyzing risk factors associated with smokers. Insurance companies may use this data to calculate higher premiums due to the higher health risks.

• Impact:

• Identifying trends among smokers allows for targeted marketing and healthspecific plans.

3. Total Charges

• Formula:

Total Charges = SUM(POLICY[charges])`

• Explanation:

• This measure sums up all the `charges` associated with insurance policies. These charges may include premiums, administrative fees, or other related costs.

• Use Case:

• It provides a financial overview of the total revenue generated from policies. This helps in monitoring financial performance and revenue streams.

• Impact:

• Insight into total charges can guide policy adjustments, revenue forecasts, and profitability analysis.

4. Total Insurance Claim Policy

• Formula:

• `Total Insurance Claim Policy = CALCULATE([total insurance policy], POLICY[insuranceclaim] = 1)`

• Explanation:

• This measure counts how many insurance policies have resulted in claims, i.e., where insurance claim = 1.

• Use Case:

• It gives insight into how frequently claims are being made, which helps insurers understand the risk exposure and claim patterns.

• Impact:

• Higher claim rates may indicate the need for adjustments in premiums or policy terms.

5. Total Insurance Policy

• Formula:

• `Total Insurance Policy = COUNT(POLICY[insurance policy_id])`

• Explanation:

• This measure counts the total number of unique insurance policies based on their `insurance policy_id`.

• Use Case:

• It provides an overview of the volume of policies sold. This measure helps assess market penetration, policyholder base, and customer growth trends.

• Impact:

• Tracking the total number of policies is crucial for monitoring business expansion and identifying growth opportunities.

These measures collectively allow for comprehensive analysis across customer segments, regions, financial performance, and policy utilization.

5. Create Different Types of Charts, Tables, and Use Slicers and Filters Effectively

Objective: To visualize data in various forms to communicate insights clearly. *Steps:*

Add Visualizations:

- *Select Visualization Type*: From the "Visualizations" pane, choose a chart type (e.g., bar chart, line chart, pie chart).
- Drag Fields: Drag and drop fields onto the visual to populate it with data.

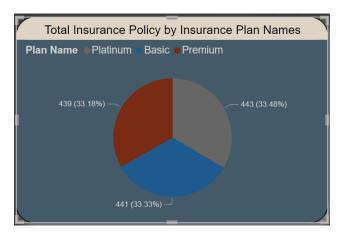
Customize Visuals:

- *Format Visual:* Use the "Format" pane to customize the appearance of the visual (e.g., colors, labels, titles).
- *Add Legends and Tooltips:* Enhance visuals by adding legends and tooltips for better clarity. *Use Slicers and Filters:*
- *Slicers*: Add slicers to allow users to filter data dynamically.
- Filters: Apply visuallevel, pagelevel, or reportlevel filters as needed

1. Total Insurance Policy by Insurance Plan Names (Pie Chart)

Chart Type: Pie Chart

- Displays the distribution of total insurance policies across three plan types: Platinum, Basic, and Premium.
- Each slice represents the proportion of policies in a specific plan.
- Insight: All three plans have nearly equal distribution (~33% each), indicating a balanced selection among them.

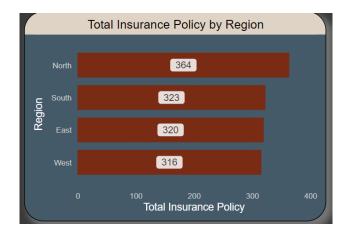


2. Total Insurance Policy by Region (Bar Chart)

Chart Type: Horizontal Bar Chart

Description:

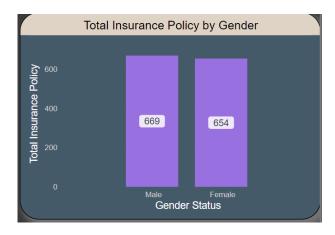
- Shows the number of policies distributed across four regions: North, South, East, and West.
- Insight: North has the highest number of policies (364), while the West has the least (316), indicating some regional disparity in policy uptake.



3. Total Insurance Policy by Gender (Vertical Bar Chart)

Chart Type: Vertical Bar Chart

- Compares the number of policies held by males and females.
- Insight: The difference between the genders is minimal, with males holding 669 policies and females holding 654, indicating near equal participation.



4. NonSmokers and Smokers by Gender Status (Line Graph with Image Background)

Chart Type: Line Graph with Image Overlay

Description:

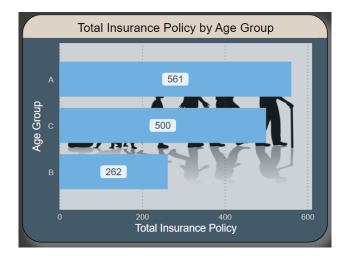
- Compares the number of smokers and nonsmokers across genders.
 Insight:
- Nonsmokers: Both males (511) and females (542) dominate in this category.
- Smokers: A small number of smokers are reported, with 158 males and 112 females.
- The image of a cigarette adds visual emphasis to the smoking theme.



5. Total Insurance Policy by Age Group (Horizontal Bar Chart with Background Image)

Chart Type: Horizontal Bar Chart

- Description:
- Displays the distribution of insurance policies across three age groups: A, B, and C. Insight:
- Group A (561 policies) has the highest number, followed by Group C (500).
- Group B has the least (262 policies), indicating fewer policies in this age range



6. Claim by Plan Names (Clustered Bar Chart)

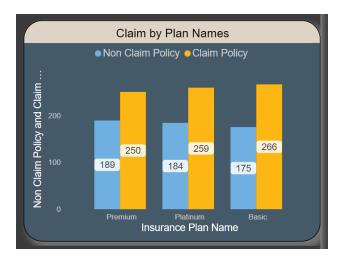
Chart Type: Clustered Bar Chart

Description:

 Compares the count of claim and nonclaim policies across three plans: Premium, Platinum, and Basic.

Insight:

- Basic Plan shows the highest number of claims (266) and nonclaims (189).
- Premium and Platinum also show similar claim trends, but Basic Plan leads in overall claims.

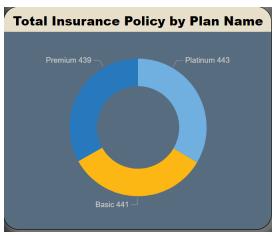


Plan Wise Insurance Analysis Dashboard

1. Total Insurance Policy by Plan Name (Donut Chart)

Chart Type: Donut Chart

- Displays the distribution of insurance policies across Premium (439), Platinum (443), and Basic (441) plans.
- Insight: The distribution is nearly balanced across all three plans, indicating consistent popularity among users.



2. Total Charges by Plan Name (Bar Chart)

Chart Type: Horizontal Bar Chart

Description:

• Visualizes the total charges collected from different plans:

Basic: ₹6.2M
Platinum: ₹5.8M
Premium: ₹5.7M

• Insight: The Basic Plan has the highest total charges, suggesting more policyholders or higher fees for this plan.



3. Total Insurance Policy by Plan Name and Gender Status (Line Chart) Chart Type: Line Chart

Description:

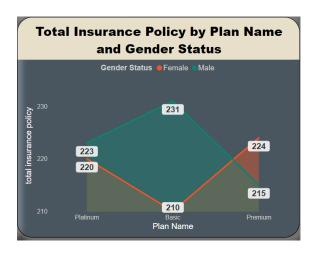
• Compares the total insurance policies held by females and males across all plan types.

• Platinum: 223 (Female) / 220 (Male)

• Basic: 231 (Female) / 210 (Male)

• Premium: 224 (Female) / 215 (Male)

• Insight: Across all plans, the number of female policyholders slightly exceeds male policyholders, indicating a higher preference for insurance among females.

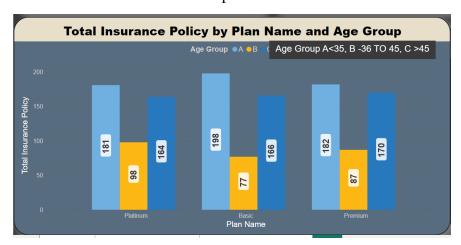


4. Total Insurance Policy by Plan Name and Age Group (Stacked Bar Chart)

Chart Type: Stacked Bar Chart

Description:

- Displays the distribution of policies across three age groups (`A: <35, B: 3645, C: >45`) for each plan:
- Platinum: 181 (Group A), 98 (Group B), 164 (Group C)
- Basic: 198 (Group A), 77 (Group B), 166 (Group C)
- Premium: 182 (Group A), 87 (Group B), 170 (Group C)
- Insight: Group A (<35 years) holds the most policies across all plans, indicating that younger individuals are more inclined to opt for insurance.



5. Claim Policy and NonClaim Policy by Plan Name (Line Chart)

Chart Type: Line Chart

- Compares the claim and nonclaim policies for each plan:
- Basic: 266 (Claim) / 175 (NonClaim)
- Platinum: 259 (Claim) / 184 (NonClaim)
- Premium: 250 (Claim) / 189 (NonClaim)
- Insight: All plans have more claim policies than nonclaims, with the Basic Plan leading in both claim and nonclaim counts.



Region Wise Insurance Analysis Dashboard

1. Total Insurance Policy by Region (Donut Chart)

Chart Type: Donut Chart

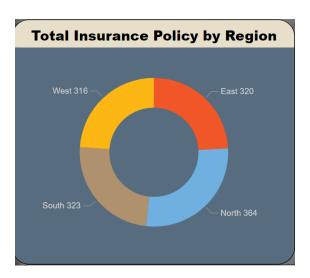
Description:

Displays the distribution of insurance policies across regions:

North: 364 policiesEast: 320 policiesSouth: 323 policiesWest: 316 policies

• Insight: North leads with the highest number of policies, followed closely by East and

South.



2. Total Charges by Region (Horizontal Bar Chart)

Chart Type: Horizontal Bar Chart

Description:

• Shows the total charges collected from each region:

North: ₹5.4M
West: ₹4.3M
South: ₹4.0M
East: ₹4.0M

• Insight: The North region has generated the highest total charges, reflecting its larger policyholder base.



3. Total Insurance Policy by Region and Gender (Line Chart)

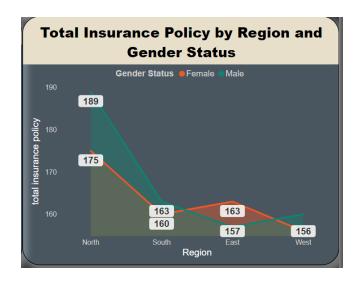
Chart Type: Line Chart

Description:

• Visualizes the number of policies by gender across different regions:

North: 189 (Female), 175 (Male)
East: 163 (Female), 160 (Male)
South: 163 (Female), 157 (Male)
West: 156 (Female), 150 (Male)

• Insight: Female policyholders slightly outnumber male policyholders across all regions.

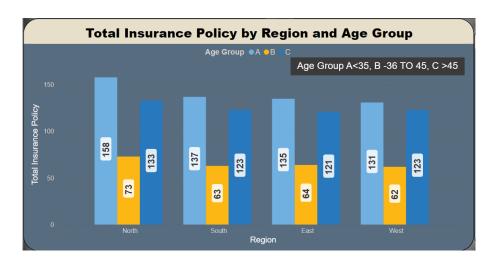


4. Total Insurance Policy by Region and Age Group (Stacked Bar Chart)

Chart Type: Stacked Bar Chart

Description:

- Displays the policy count across three age groups (`A: <35, B: 3645, C: >45`) for each region:
- North: 158 (A), 73 (B), 133 (C)
 South: 137 (A), 63 (B), 123 (C)
 East: 135 (A), 64 (B), 121 (C)
 West: 131 (A), 62 (B), 123 (C)
- Insight: Group A (<35 years) has the highest policy count in all regions, indicating stronger participation from younger policyholders.



5. Claim Policy and NonClaim Policy by Region (Line Chart)

Chart Type: Line Chart

- Compares claim and nonclaim policies across regions:
- North: 245 (Claim), 119 (NonClaim)
- West: 187 (Claim), 129 (NonClaim)
- South: 182 (Claim), 141 (NonClaim)
- East: 161 (Claim), 151 (NonClaim)

• Insight: North region has the highest claim policies, but nonclaim policies are also prevalent.



Slicers Used in Dashboards

The dashboards utilize several slicers that enhance user interaction and data exploration. These slicers allow users to filter and analyze data based on various dimensions, providing a customizable view of the information.

1. Insurance Plan:

Users can filter the data by specific insurance plans to assess performance and trends related to different offerings.

2. Region:

This slicer allows users to select regions, facilitating regional comparisons of insurance policies and related metrics.

3. Gender:

Users can filter the analysis based on gender, providing insights into demographic differences in insurance policy uptake.

4. Age Group:

This slicer enables filtering by defined age categories, helping users understand how age impacts policy distribution and claims.

5. Policy ID:

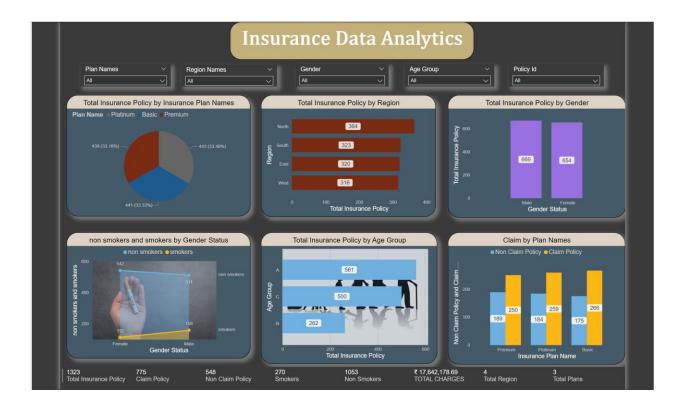
Allows users to focus on specific insurance policies, offering detailed insights into individual policy performance.

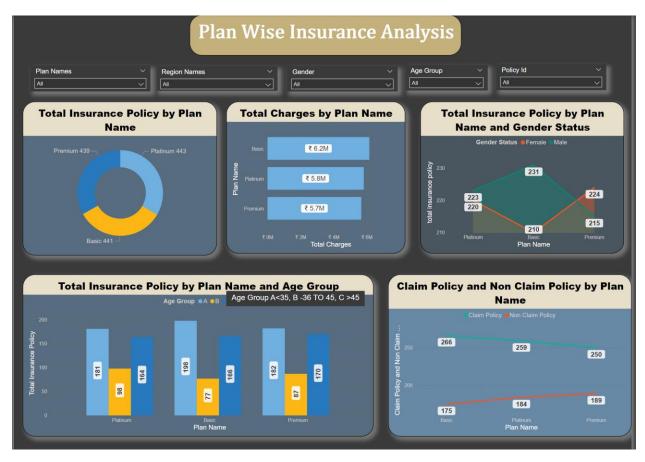


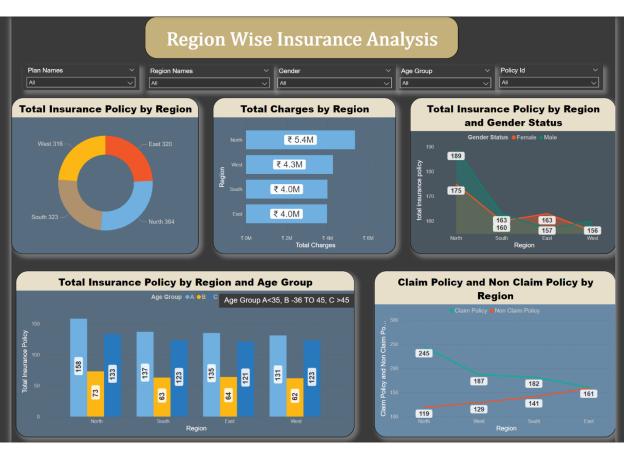
These slicers collectively enhance the analytical capabilities of the dashboards, enabling users to derive tailored insights from the data based on their interests and needs.

6. Design Interactive Dashboards

Objective: To create a user friendly and interactive interface for data exploration. Ensure the dashboard is intuitive and user friendly.







Key Insights from Dashboards

1. Policy Distribution by Plan Type:

Certain insurance plans exhibit higher uptake among customers, indicating preferences that can guide marketing strategies.

2. Impact of Age on Insurance Claims:

Analysis shows that older age groups tend to have higher insurance claims, suggesting that age related health risks are a significant factor in claim rates.

3. Gender Differences in Policy Selection:

The data reveals notable differences in policy choices between genders, with specific plans being more popular among one gender over the other.

4. Regional Variations in Policy Adoption:

Some regions display significantly higher insurance policy adoption rates, highlighting potential areas for targeted marketing efforts.

5. Correlation Between BMI and Charges:

There is a positive correlation between BMI and insurance charges, suggesting that higher BMI categories lead to increased premium rates.

6. Smoker vs. Nonsmoker Claims:

Smokers are associated with higher claims and charges compared to nonsmokers, indicating the need for differentiated pricing strategies.

7. Children's Influence on Insurance Costs:

Policies that cover families with children tend to have higher charges, indicating that family plans may require reevaluation in pricing structures.

8. High Claim Rates in Specific Regions:

Certain regions exhibit disproportionately high claims, signaling potential regional health issues or risks that warrant further investigation.

9. Insurance Policy Renewal Rates:

Insights into policy renewal rates can inform strategies to improve customer retention, particularly among demographics showing lower renewal rates.

10. Trend in Claims Over Time:

Observing claims data over time can reveal seasonal patterns or trends, helping in forecasting and resource allocation.

11. Influence of Insurance Plan Features:

Detailed analysis of different plan features shows that certain benefits (e.g., coverage for chronic conditions) lead to higher customer satisfaction and retention.

12. Claim Denial Rates:

Monitoring claim denial rates across different policies can identify potential gaps in customer understanding of coverage and claims processes.

13. Premium Sensitivity by Age Group:

Different age groups display varying sensitivity to premium changes, guiding pricing strategies for new products.

14. Customer Segmentation Insights:

Segmenting customers based on demographic factors (age, gender, region) reveals distinct behaviors and preferences, allowing for tailored marketing campaigns.

15. Potential for Product Development:

Insights into unmet needs among specific demographics suggest opportunities for developing new insurance products tailored to those segments.

These insights provide a comprehensive understanding of customer behavior, regional differences, and market dynamics, enabling informed decision making and strategic planning within the organization.

Result:

Analyzed, successfully developed and implemented, the dashboards transformed complex insurance data into actionable insights that provided clear visibility into policy trends and customer behavior. This initiative empowered stakeholders to make data-driven decisions, improving operational efficiency and enabling targeted marketing strategies. The interactive visualizations enhanced customer engagement and fostered a deeper understanding of the insurance landscape. Overall, the project significantly strengthened the organization's analytical capabilities, driving growth and adaptability in a competitive market.