

Insurance Customer Insights & Retention Analytics

Problem Statement:

A leading insurance company wants to increase sales of **motor 🚗** and **health 🩺** insurance policies by identifying a target customer base with high purchase or renewal potential.

The company's customer data is stored across multiple tables with inconsistent formats, missing values, and no unified view. This makes it difficult to:

- ⌚ Identify **high-value customers** with strong financial profiles.
- 🕒 Track **policies nearing expiry** to prevent churn.
- 📈 Analyze **demographics and premium patterns** for targeted marketing.

This project integrates data from multiple sources using **SQL** and visualizes insights in **Power BI** to support **data-driven marketing and retention strategies**.

❖ Data Overview:

The synthetic dataset contains the following tables:

1. 🧑 Personal Information Table

- **Description:** Personal details with standardized gender and corrected email IDs.

Transformations Applied:

- Standardized gender values (M → Male, F → Female).
- Filled missing email_id with name-based placeholders (e.g., JohnDoe@example.com).
- Converted dob from text to DATE format.

○ Columns:

PAN_no, name, gender, occupation, mobile, dob, email_id

```
14 •  select * from personal_data;
```

Result Grid | Filter Rows: Export: Wrap Cell Content: Fetch rows:

PAN_no	name	gender	occupation	mobile	dob	email_id
oiVg1043R	Jordan Smith	Male	Unknown	3338679832	1976-02-10	christinahenderson@example.net
cbfn8196o	Kimberly Johnson	Male	Unknown	(512)468-1	2002-06-20	msavage@example.com
mTPS8908I	Dana Jackson	Male	Doctor	657-344-45	1983-06-28	jennifer88@example.com
rZaW6379Z	Robin Ford	Male	Doctor	8328678879	1984-02-01	michael73@example.org
jnWv2654g	Matthew Le	Male	Doctor	313-201-44	1982-10-20	bking@example.net
wMqZ1615c	Chris Gordon	Female	Engineer	686-289-06	1986-05-15	smithanthony@example.net

2. Demographic Information Table

- **Description:** Customer demographic details after joining with city_data and standardizing pincodes.

Transformations Applied:

- Removed duplicate state column before join.
- Joined with city_data to add correct state and city name.
- Standardized pincode based on city rules.
- Converted all pincodes to valid 6-digit integers.

○ Columns:

PAN_no, city_key, state, city_name, tier, pincode

```
5 • select * from demographic_data;
```

6

Result Grid | Filter Rows: Export: Wrap Cell Content: Fetch rows

PAN_no	city_key	state	city_name	tier	pincode
oiVg1043R	C003	Gujarat	Ahmedabad	2	380335
cbf81960	C002	Maharashtra	Pune	2	411000
mTPS8908I	C003	Gujarat	Ahmedabad	2	380453
rZaW6379Z	C006	Delhi	Delhi	1	110082
jNwv2654g	C003	Gujarat	Ahmedabad	2	380584
wMqZ1615c	C005	Tamil Nadu	Chennai	2	600003

3. City Information Table

- **Description:** Contains final city information with added state column. Used for mapping customer location.

Transformations Applied:

- Added state column based on city_name.
- Mapped cities like Mumbai/Pune → Maharashtra, Chennai → Tamil Nadu, etc.

○ Columns:

city_key, city_name, tier, state

```
18 • select * from city_data;
```

19

Result Grid | Filter Rows: Export: Wrap Cell Content: Fetch rows

city_key	city_name	tier	state
C001	Mumbai	1	Maharashtra
C002	Pune	2	Maharashtra
C003	Ahmedabad	2	Gujarat
C004	Bangalore	1	Karnataka
C005	Chennai	2	Tamil Nadu
C006	Delhi	1	Delhi

4. Financial Information Table

- **Description:** Customer financial profile with classified card type based on income and CIBIL score.

Transformations Applied:

- Added card_type_new column.
- Classified Credit card if income ≥ 600000 AND cibil_score > 700 , else Debit.
- Dropped old card_type column and renamed new one.
- **Columns:**
PAN_no, income, cibil_score, bank_account_number, card_type

```
8 •   select * from financial_data;
```

Result Grid | Filter Rows: Export: Wrap Cell Content: Fetch rows

	PAN_no	income	cibil_score	bank_account_number	card_type
▶	oiVg1043R	594125	781	CAVB12049850531356	Debit
	cbfn8196o	513688	717	QLRV21139022478227	Debit
	mTPS8908I	651762	643	QGIJ21175115184492	Debit
	rZaW6379Z	895776	773	JTVB80409951177443	Credit
	jnWv2654g	860832	701	DNQI18701852384564	Credit
	fJsO7816N	621503	708	EBMA03986864312886	Credit

5. Insurance History Table

- **Description:** Policy details with corrected date formats for start and end dates.

Transformations Applied:

- Converted insurance_start_date and insurance_end_date from text to DATE format (DD-MM-YYYY → YYYY-MM-DD).
- Validated dates to ensure non-null values for existing policies.

○ **Columns:**

PAN_no, type_of_insurance, premium, payment_frequency, insurance_start_date, insurance_end_date

```
11 •   select * from insurance_data;
```

Result Grid | Filter Rows: Export: Wrap Cell Content: Fetch rows

	PAN_no	type_of_insurance	premium	payment_frequency	insurance_start_date	insurance_end_date
▶	CVuj0033l	Motor	16682	Yearly	2022-10-11	2025-10-09
	nNxc7268X	Motor	17146	Yearly	2024-05-15	2026-02-01
	xqpU6471O	Motor	11658	Yearly	2023-03-12	2026-08-12
	xPPM3343E	Health	29446	Monthly	2024-05-27	2026-07-01
	Jexw2781n	Motor	37941	Monthly	2023-10-16	2027-05-07
	yjsG7306e	Health	26206	Yearly	2022-12-09	2026-06-05

Project Methodology:

Step 1: Data Cleaning & Integration (Using SQL)

- Mapped city names to states in city_data.
- Standardized pin codes according to city rules.
- Normalized gender values (M → Male, F → Female).
- Classified payment card type using income and CIBIL score.
- Filled missing email IDs with placeholders.
- Converted text dates to date format.
- Merged all cleaned tables into a single **final_information** dataset for analysis.

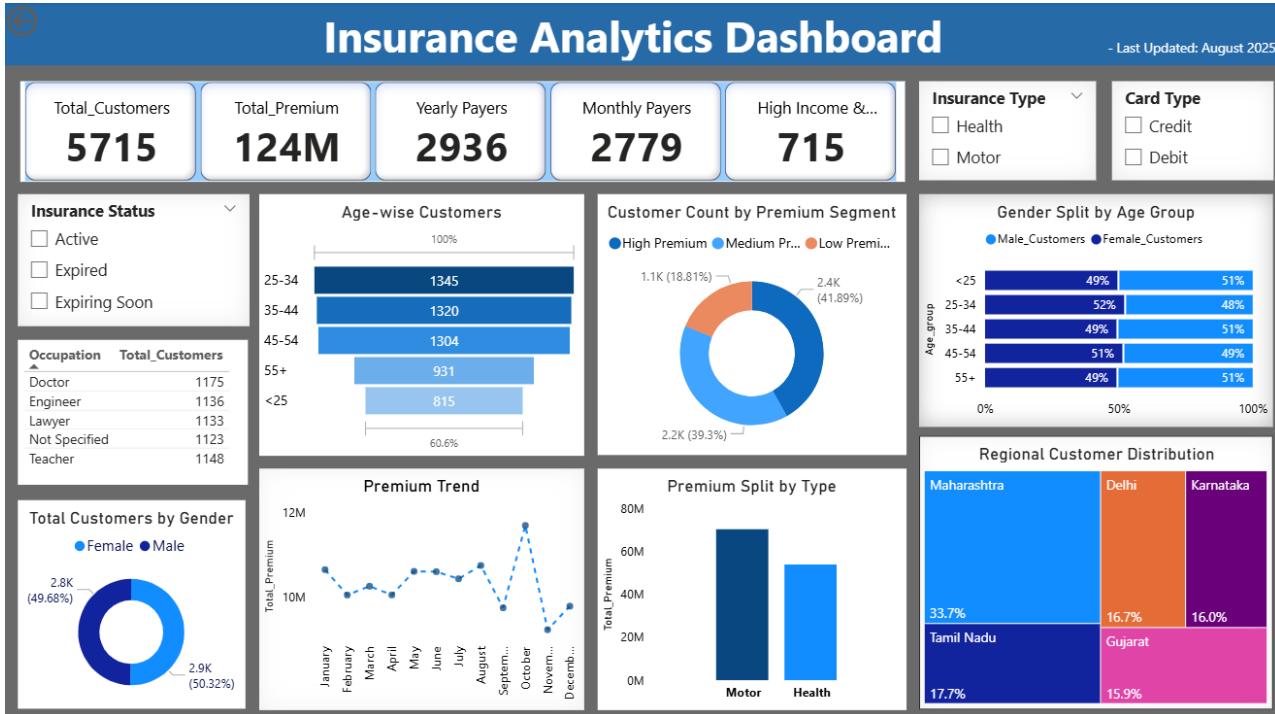
Step 2: Data Import & Visualization Setup (Using Power BI)

- Imported the cleaned final_information table into Power BI.
- Created calculated columns for **age group** and **premium category**.
- Defined DAX measures for key metrics such as **Total Premium**, **High-Value Customer Count**, and **Expiring Soon %**.

Step 3: Dashboard Creation (Using Power BI)

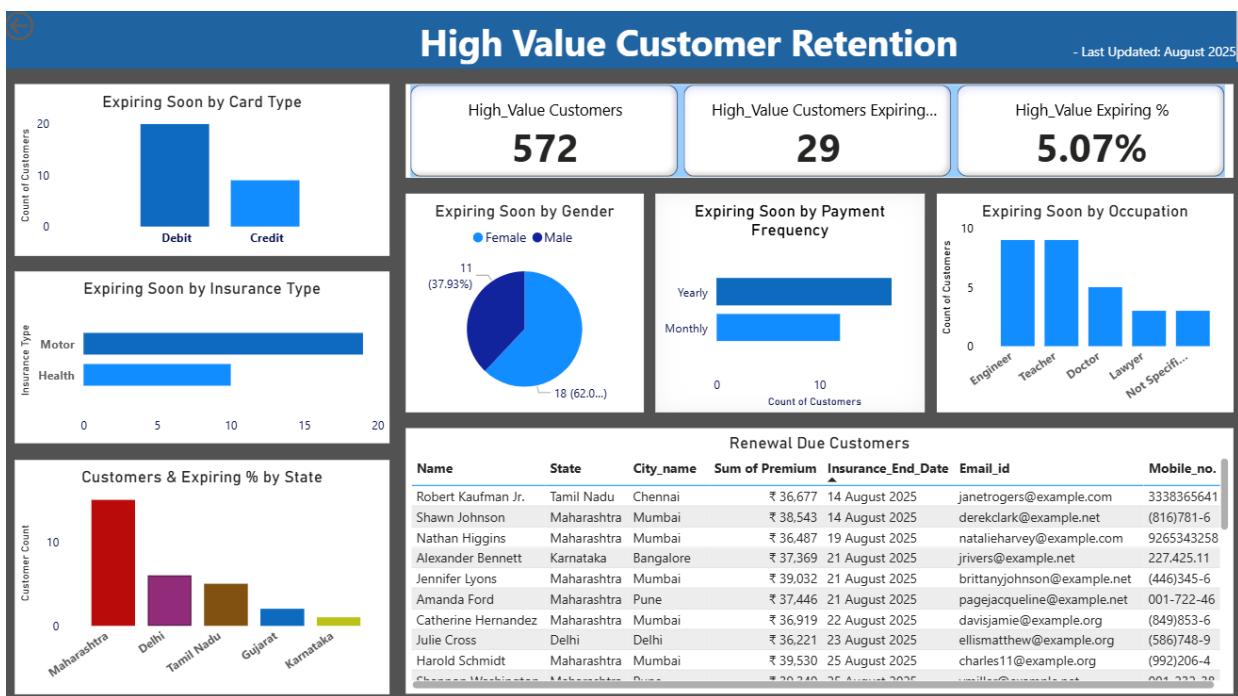
Dashboard 1 – Overall Customer Insights

- **Total Customers:** 5,715
- **Total Premium:** ₹124M
- **Yearly Payers:** 2,936
- **Monthly Payers:** 2,779
- **High Income & High CIBIL Customers:** 715
- Age Group: **25-44 years (~72%)** dominate.
- Gender Split: **Male (50.32%)**, Female (49.68%).
- Premium Segmentation: High (18.81%), Medium (39.3%), Low (41.89%).
- Top States: Maharashtra (33.7%), Tamil Nadu (17.7%), Delhi (16.7%).
- Motor Insurance generates more revenue than Health Insurance



📌 Dashboard 2 – High Value Customer Retention

- High Value Customers:** 572
- Expiring Soon:** 29 (5.07% of high-value customers).
- Expiring Soon Profile:** **62% Male**, mostly yearly payers.
- State with Most At-Risk Customers:** Maharashtra.
- Top Occupations at Risk:** Engineer, Teacher, Doctor.



Output and Business Value:

- Identified **572 high-value customers** for targeted retention campaigns.
 - Provided **124M premium revenue visibility** with segment-wise breakdown.
 - Pinpointed **29 high-value customers at risk of churn**.
 - Highlighted top-performing states for marketing focus.
 - Enabled sales teams to target the most profitable and at-risk customers efficiently.
-

Tools Used:

Tool	Purpose
SQL 	Data cleaning, integration, transformation
Excel 	Validation and exploratory analysis
Power BI 	Data visualization and dashboard creation

Next Steps:

- Launch **renewal campaigns** for expiring high-value customers.
 - Explore **cross-sell opportunities** for customers with high CIBIL and income but no certain insurance types.
 - Automate **monthly dashboard refresh** with updated customer data.
 - Provide **sales team training** on using dashboard filters for lead targeting.
-

Conclusion:

This project transformed raw, inconsistent data into a clean, analysis-ready dataset and developed dashboards that provide **clear, actionable insights** for the insurance business.

Key Highlights for the Insurance Company:

- **Total Customers:** 5,715
- **High-Value Customers:** 572
- **High-Value Customers Expiring Soon:** 29 (5.07%)
- **Largest Age Segment:** 25-44 years (~72%)
- **Top State by Customers:** Maharashtra (33.7%)
- **Most Profitable Insurance Type:** Motor Insurance

These findings can directly improve **customer retention, marketing focus, and revenue growth**.