Insurance Analytics Platform: From Data to Decisions

Final Project Presentation

Elena Nurullina – July 11, 2025

Project Overview

Over the past few days, I've built an end-to-end insurance analytics platform using Azure Storage Accounts/Containers, Azure Databricks, Azure Data Factory, and Power BI.

Outcome: End-to-end data platform from raw to strategic insights:
- Environment Setup and Data Validation
- Customer Risk Profiling
- Customer Lifetime Premium Values and Retention Modeling
- Executive Dashboard

The project includes 4 Databricks notebooks, a fully orchestrated ADF pipeline, and a business-focused Power BI dashboard.

Databricks 4-Notebook Architecture: 4 Stages of Analytics

A diagram showing 4 notebooks in sequence.

The project is structured into 4 modular notebooks:

- **Notebook 0:** Environment Setup & Validation – load and validate 6 insurance datasets
- **Notebook 1:** Risk Profiling – assign risk scores and customer segmentation
- **Notebook 2:** CLPV & Retention Modeling – lifetime value modeling and renewal prediction
- **Notebook 3:** Executive Dashboard Executive PBI – export BI-ready data and KPIs

This modularity supports maintainability, reusability, and clear separation of concerns.

Notebook 0 - Environment Setup
Foundation First: Environment & Data Validation

Loaded and validated 6 datasets from `/mnt/coursedata/`
Ensured referential integrity (e.g. customer_id across tables)
Created optimized temporary views for analysis / for efficient querying
Baseline metrics: total customers, active policies, claim rate
This ensures reliable data for all downstream analysis

Predictive Insights: Risk, CLPV, and Retention
Notebook 1 - Risk Profiling

Risk Profiling:

- Calculated customer risk scores for 15,000 customers using policy, claims, and payment history
- Segmented by risk category and behavior patterns, as segmented customers into Low/Medium/High risk
- Identified cross-selling opportunities as identified seasonal claim and payment behavior

Notebook 2 - CLPV & Retention

CLPV Modeling:

- Modeled lifetime premium value per customer as Modeled Customer Lifetime Premium Value (CLPV)
- Created renewal probability scores and fraud indicators
- And predicted renewal probabilities & detected fraud risks
- Developed retention and pricing strategies as outputs used to prioritize retention strategies

From Analysis to Action: Executive Insights
Notebook 3 - BI Dashboard Prep Export & KPIs

Calculated executive KPIs: Total Premiums, Loss Ratio, Retention Rate
Prioritized high-value customers for marketing (Flagged high-value customers for targeting)
Generated Power BI-ready exports with clean formats
Prepared Power BI-ready export tables as csv downloaded files
Built a performance monitoring framework to track business health

Automation with Azure Data Factory
Azure Data Factory Pipeline

Built an ADF pipeline:

4 Databricks notebook activities, 2 web alerts: Webhooks (Success/Failure)
Set up sequential dependencies: 0 → 1 → 2 → 3

Added webhooks for success/failure notifications using `httpbin.org`
Scheduled to run daily at 4:00 AM with retries and monitoring

Power BI Dashboard

Page 1: KPIs, trends, policy mix, geo as Executive Insurance Overview
4 KPI cards: Total Premiums, Loss Ratio, Customer Count, Retention Rate
Line Chart: Date vs Total Premiums
Pie Chart: policy_type by Total Premiums

Page 2: Customer Analytics having Customer CLPV, risk matrix
Column chart: X = customer_lifetime_premium_value, Y = count of customers
Scatter: X: composite_risk_score Y: customer_lifetime_premium_value Legend: risk_category Table: Columns: risk_category, value_segment Measures: Count of customers, Avg CLPV

Page 3: Claims and high-risk analysis
Line chart: Claims over time (claim_date by month, Y: claim_amount)
Table: High-risk customers only Filter: risk_category = "High"
Bar chart: policy_type vs SUM(claim_amount)

Page 4: Business Recommendations with action matrix & top customers, KPI cards

Page 1: Executive Insurance Overview

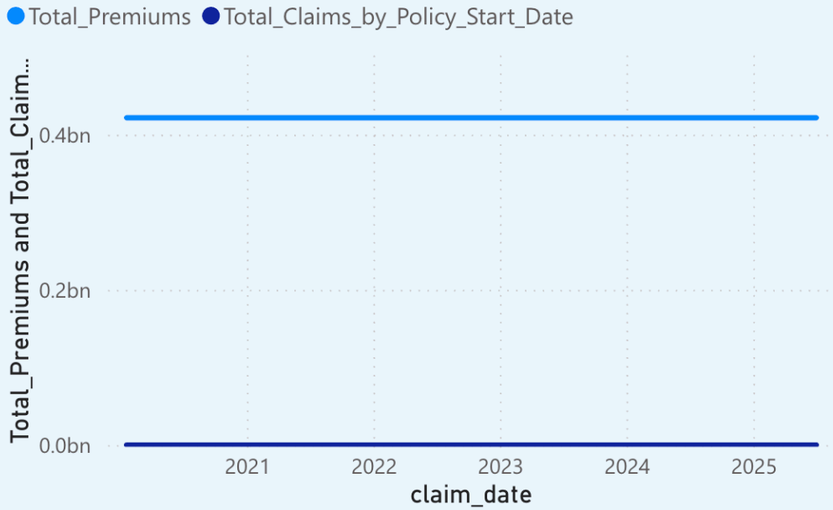
421.65M	297.03	2918	79.49
Total_Premiums	Loss_Ratio	Customer_Count	Retention_Rate

Total_Premiums, Loss_Ratio and Customer_Count by satisfaction_resolution_rate

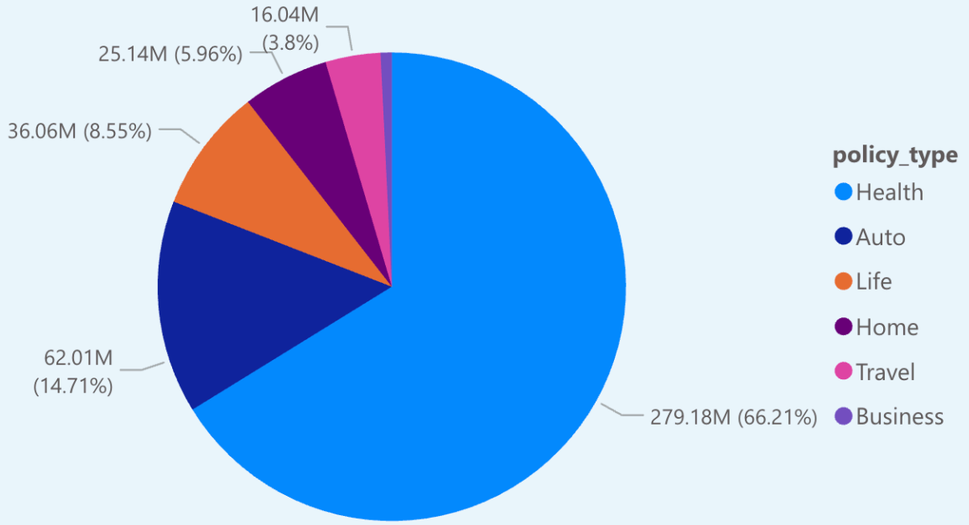
421.65M

Goal: 297, 2918

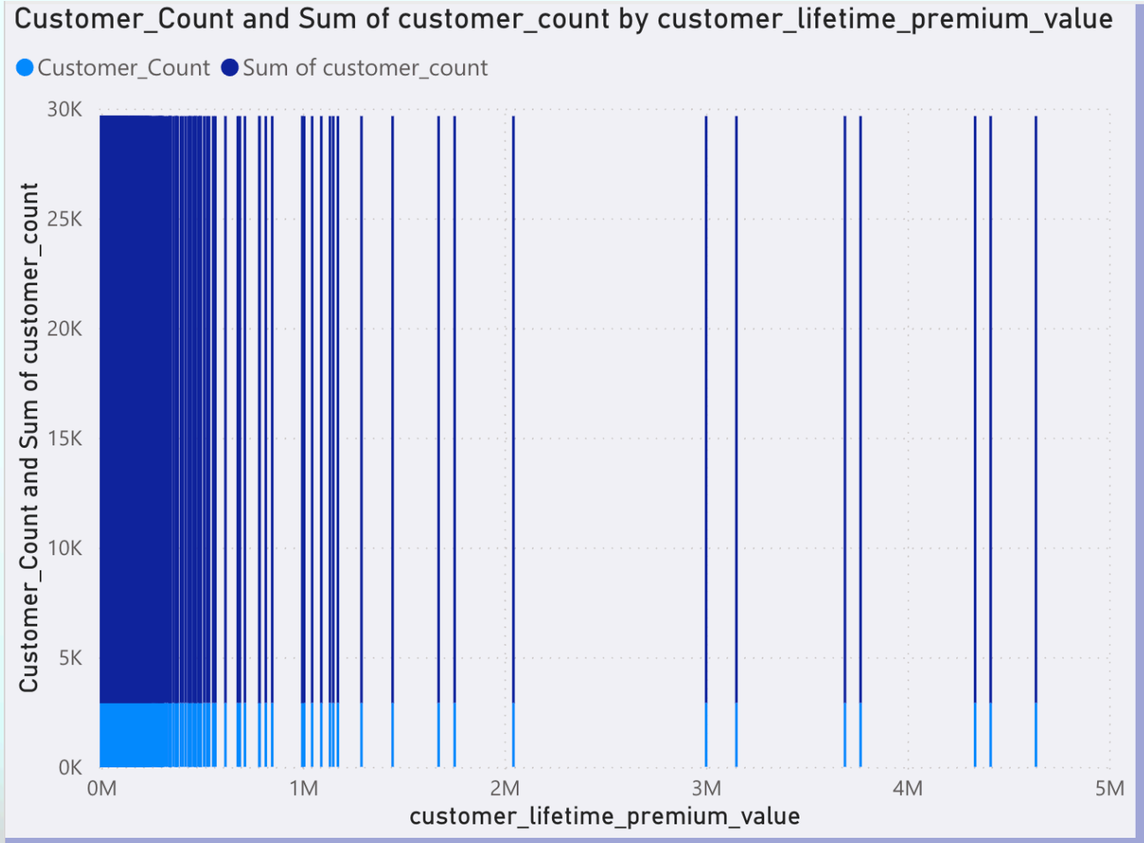
Total_Premiums and Total_Claims_by_Policy_Start_Date by claim_date

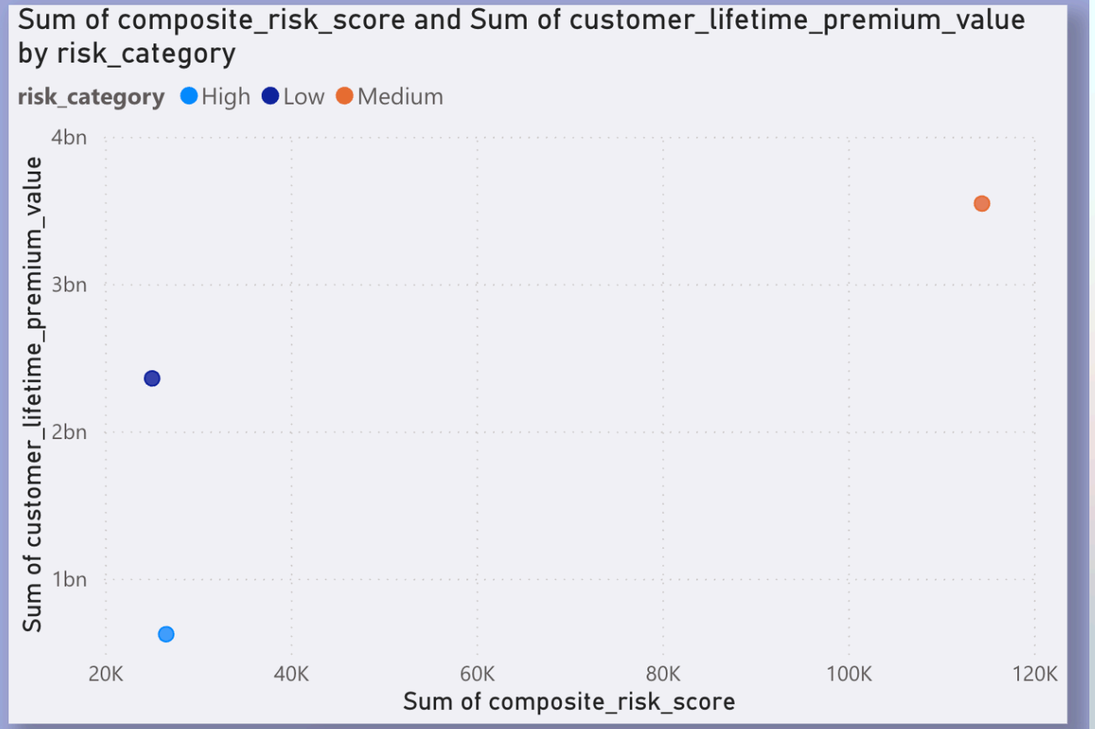


Sum of total_premium by policy_type

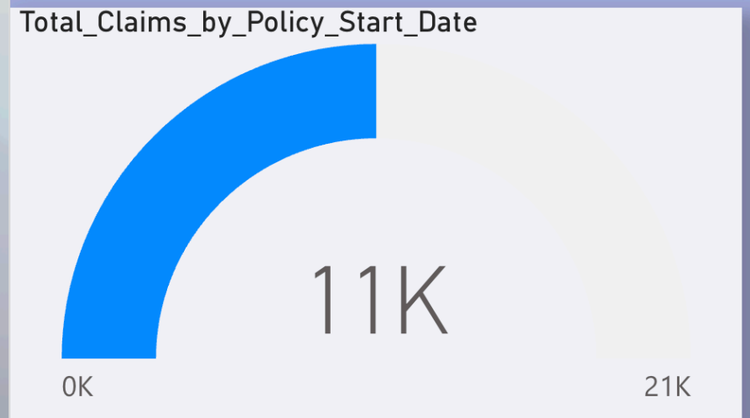


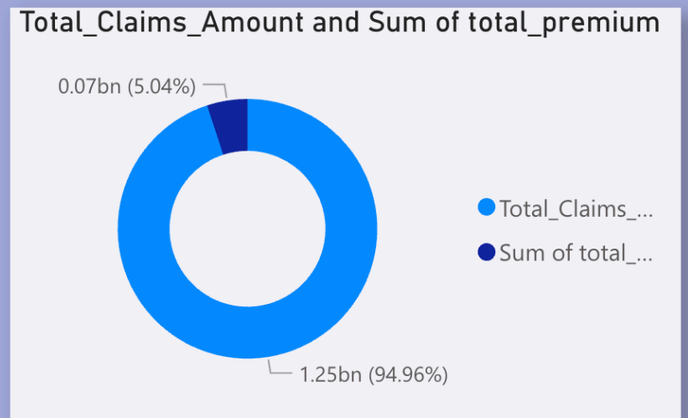
Page 2: Customer Analytics



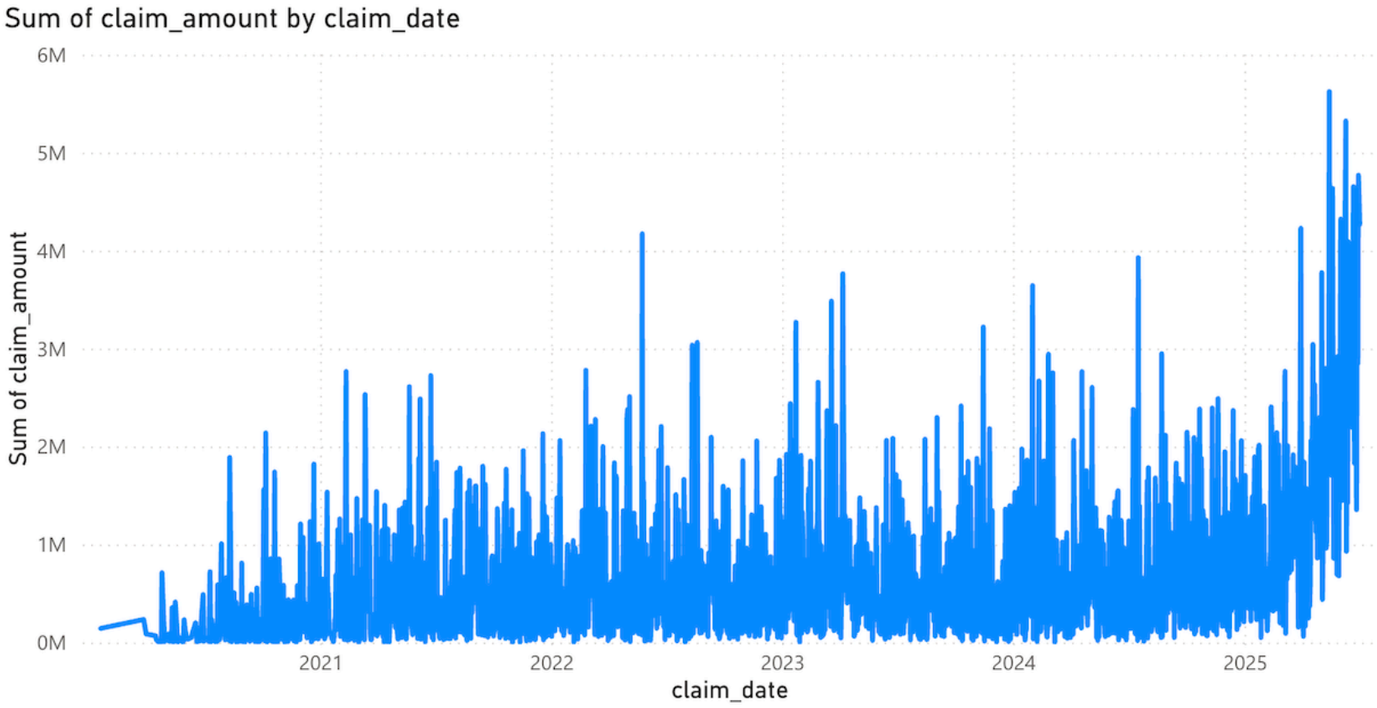


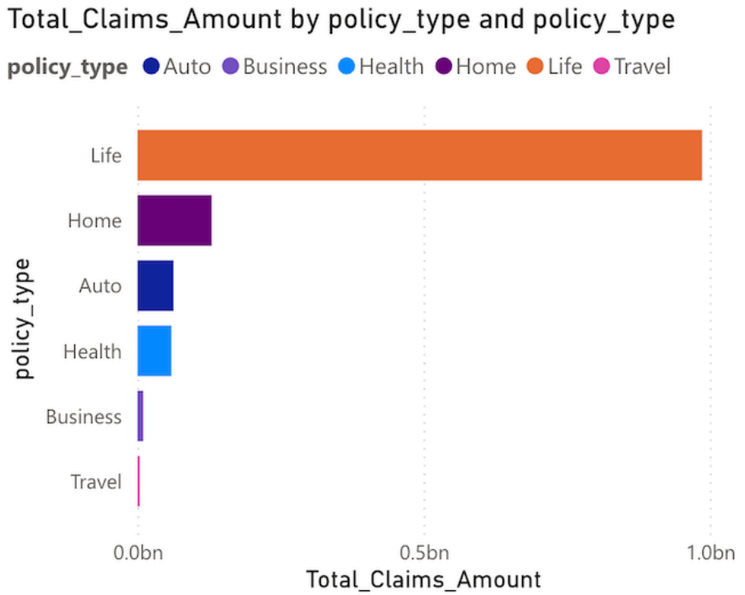
risk_category	value_segment	Customer_Count	Average_CLPV
High	High Value	2918	292,520.56
High	Low Value	2918	292,520.56
High	Medium Value	2918	292,520.56
Low	High Value	2918	292,520.56
Low	Low Value	2918	292,520.56
Low	Medium Value	2918	292,520.56
Medium	High Value	2918	292,520.56
Medium	Low Value	2918	292,520.56
Medium	Medium Value	2918	292,520.56
Total		**2918**	**292,520.56**



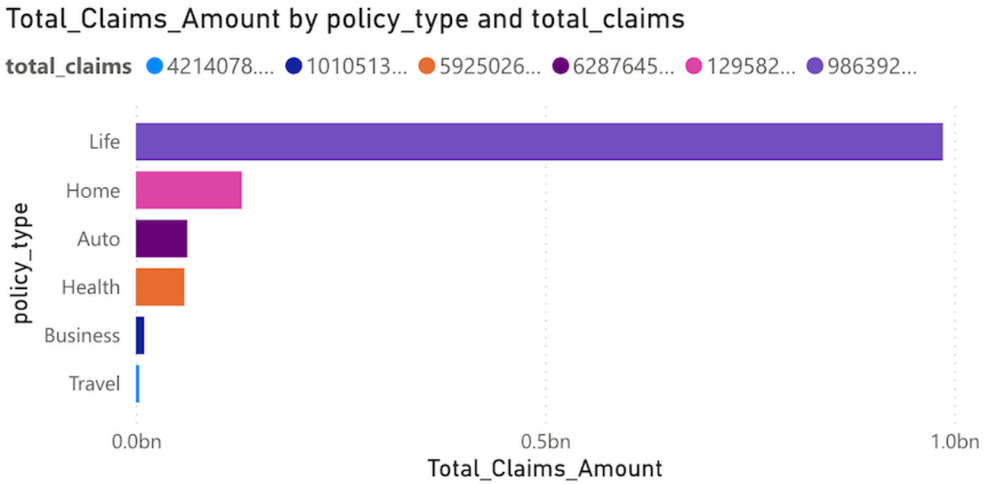


Page 3: Claims and high-risk analysis





risk_category	High-Risk_Customer_Count	first_name	last_name	customer_id	Sum of total_premium
High	1	Adam	Garcia	CUST005293	100,847.10
High	1	Adrian	Boyle	CUST005900	46,361.77
High	1	Adrienne	Martinez	CUST012000	92,098.72
High	1	Alexander	Harris	CUST007087	110,680.48
High	1	Alexander	Kim	CUST005514	102,591.38
High	1	Alexander	Lee	CUST014015	78,434.84
High	1	Alison	Oneal	CUST013137	102,289.85
High	1	Allen	Salazar	CUST004724	63,431.05
High	1	Amanda	Christian	CUST010664	28,481.81
High	1	Amanda	Wang	CUST011469	76,015.93
Total	**397**				**163,667,102.73**



Filters

Page 4: Business Recommendations with action matrix & top customers, KPI cards

Line chart:

risk_category	High Value	Low Value	Medium Value	**Total**
High	356	740	355	**1451**
Low	356	2799	574	**3729**
Medium	1480	6298	2042	**9820**
Total	**2192**	**9837**	**2971**	**15000**

first_name	last_name	Sum of customer_lifetime_premium_value
Danny	Hall	4,637,516.57
William	Obrien	4,413,970.56
Edward	Williams	4,335,715.60
Tracy	Randolph	4,284,636.15
Gary	Mckinney	3,767,354.82
Kathleen	Durham	3,689,849.76
David	Potter	3,151,383.85
Heather	Flynn	3,001,259.25
Mary	Bray	2,798,376.92
Michael	Schultz	2,045,174.27
Total		**178,185,628.61**

Final Deliverables

- 4 Databricks Notebooks
- Azure DataFactory Pipeline JSON + Screenshots
- Power BI .pbix + dashboard visuals
- Presentation PDF and summary

Conclusion & Next Steps

- Built scalable architecture and BI insights
- Next: Real-time scoring & cloud deployment

Challenges & Achievements

Challenge: Importing from Databricks to Power BI
Solved via CSV exports from Notebooks 00 & 03
Learned the value of pre-cleaned BI outputs
Future Idea: Use Azure Data Lake or a direct connector to import CSV from my Mac

Power BI:
Issue: Error due to multiple relationship paths between Calendar, Claims, and Policies tables (e.g., Calendar → Policies → Claims vs Calendar → Claims).
Resolution: Used USERELATIONSHIP() inside DAX measures to activate specific relationships without breaking the data model.

Enjoyed the process – full data-to-value cycle!
Enjoyed Most
- Working with complex relationships, ambiguous joins, and missing fields helped me sharpen my problem-solving skills.
- I enjoyed turning raw data into clear, insightful dashboards that tell a business story.
- I learned how to write advanced DAX for business logic.

Thank You!

Q&A ?