Data-Driven Business Solutions: A Portfolio of Case Studies

Application of Advanced Analytics from my MSc in Big Data Analytics

Syed Taizeem

Linked.com/in/syed-taizeem-hamdani

Bridging the Gap Between Data and Decisions

▪This portfolio showcases four projects completed during my MSc in Big Data Analytics. Each case study follows a consistent framework: defining a business problem, applying a rigorous analytical method, and delivering a clear, actionable recommendation.

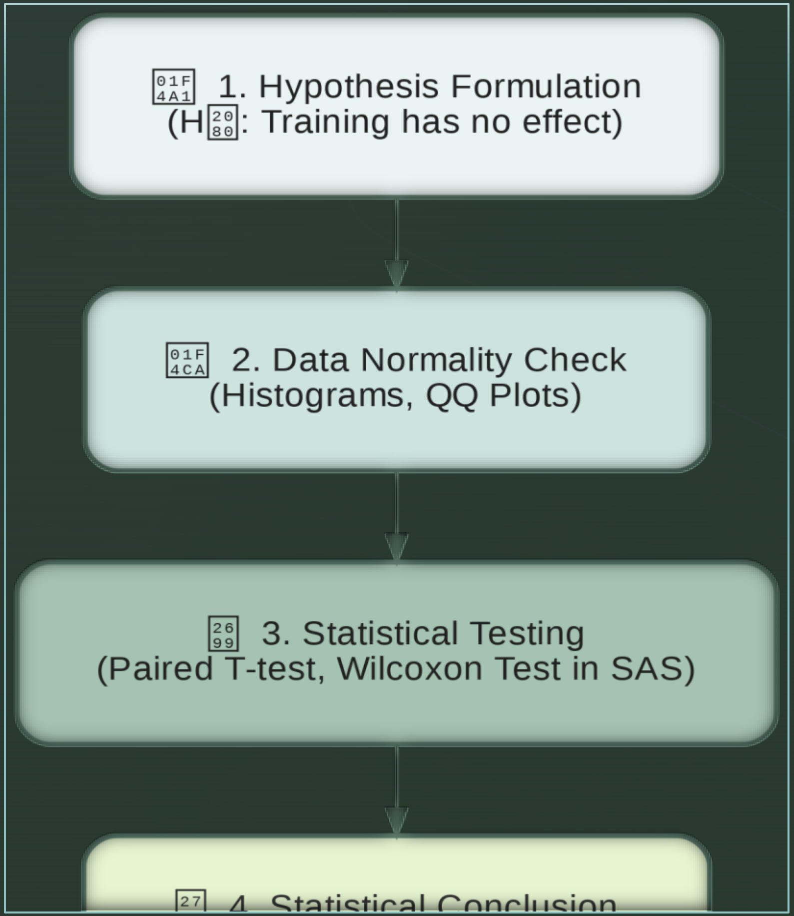
- Project 1: Retail - Sales Training Effectiveness
- Project 2: Real Estate - Investment Strategy
- Project 3: Insurance - Risk Modeling
- Project 4: Finance - Loan Payment Analysis

Retail - Sales Training Effectiveness

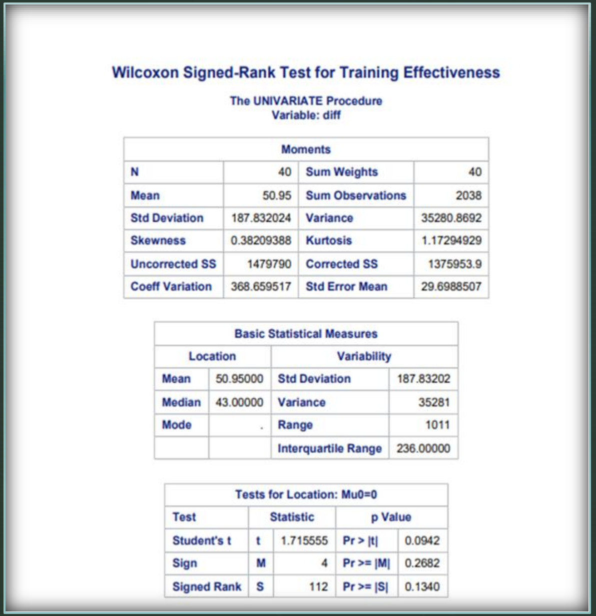
The Problem: A UK department store piloted a new training course for its cosmetics staff but was unsure if it actually improved sales performance.

The Objective: My task was to statistically determine if the training had a significant impact on sales, providing a data-driven "Go / No-Go" recommendation for a costly, company-wide rollout.

Our Data-Driven Method: Hypothesis Testing



- Both tests yielded p-values (0.094 and 0.134) greater than the 0.05 significance level. This means we cannot reject the null hypothesis; the observed sales increase is not statistically significant.



Wilcoxon Signed-Rank Test for Training Effectiveness

The UNIVARIATE Procedure
Variable: diff

Moments			
N	40	Sum Weights	40
Mean	50.95	Sum Observations	2038
Std Deviation	187.832024	Variance	35280.8692
Skewness	0.38209388	Kurtosis	1.17294929
Uncorrected SS	1479790	Corrected SS	1375953.9
Coeff Variation	368.659517	Std Error Mean	29.6988507

Basic Statistical Measures			
Location		Variability	
Mean	50.95000	Std Deviation	187.83202
Median	43.00000	Variance	35281
Mode	.	Range	1011
		Interquartile Range	236.00000

Tests for Location: Mu0=0				
Test	Statistic		p Value	
Student's t	t	1.715555	Pr > |t|	0.0942
Sign	M	4	Pr >= |M|	0.2682
Signed Rank	S	112	Pr >= |S|	0.1340

Key Finding: A Positive Trend, But Not Statistically Significant

📈 A Small Positive Effect:

- Trained salespersons showed a mean sales increase of £50.95. However, the results varied widely across individuals.

❌ Lacks Statistical Proof

- The high p-value and a 95% confidence interval that included £0 means the training's true impact is uncertain. It could range from a slight loss to a modest gain.

Business Impact & Final Recommendation

Our Recommendation:

- "Do not implement the training course chain-wide based on this pilot study."

Business Impact:

- ✅ Prevented Unnecessary Spending: Saved the company from investing significant capital into a program with no proven ROI.
- 🎯 Informed Future Strategy: Advised further investigation into other factors driving sales performance before re-investing in training.
- 💰 Clear Financial Guidance: Provided a definitive, data-backed answer to a critical business question.

Real Estate - Investment Strategy

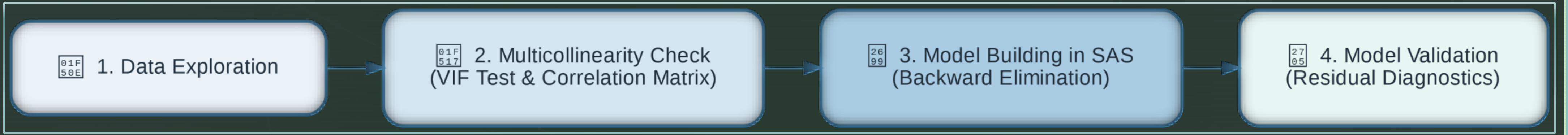
The Problem

- A property group's method for assessing new investments was slow, expensive, and relied on subjective surveys, limiting their ability to expand effectively.

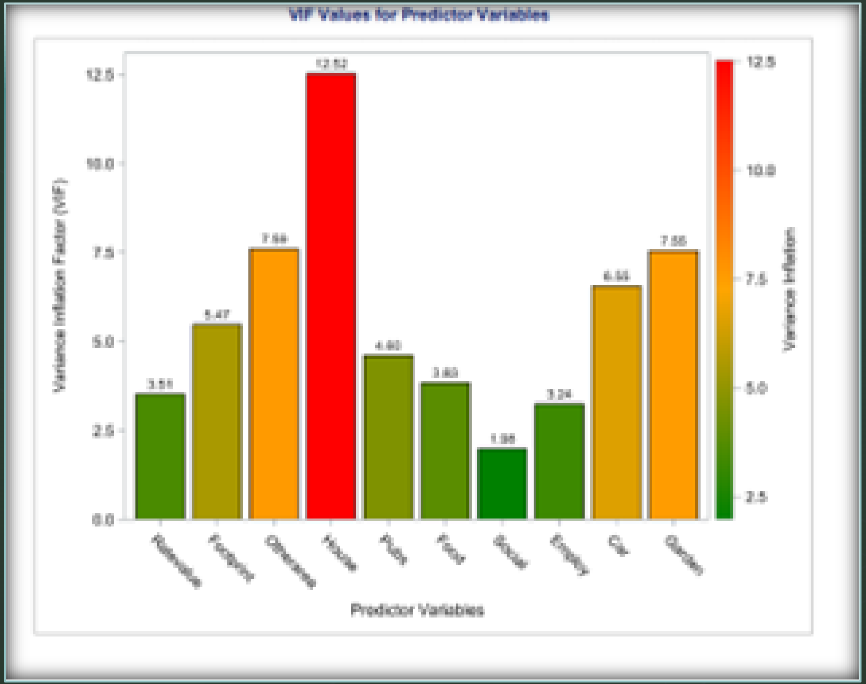
The Objective

- Our goal was to build a predictive regression model using readily available data to identify the key drivers of property income, enabling a faster, more strategic acquisition process.

Our Data-Driven Method: Multiple Regression Modeling



- The VIF test revealed severe multicollinearity (VIF > 5) in 'House' and 'Otherarea' variables. Removing them was crucial for creating a reliable and stable final model.



Key Findings: The 3 Drivers of Property Income

01

Rateable Value: The strongest predictor. For every £1k increase in rateable value, weekly income rises by £850.

02

Local Unemployment: A critical risk factor. Every 1% increase in unemployment reduces weekly income by £120.

03

Outdoor Space: A key customer amenity. Every additional square meter of garden space adds £40 to weekly income.

Business Impact & Final Recommendation

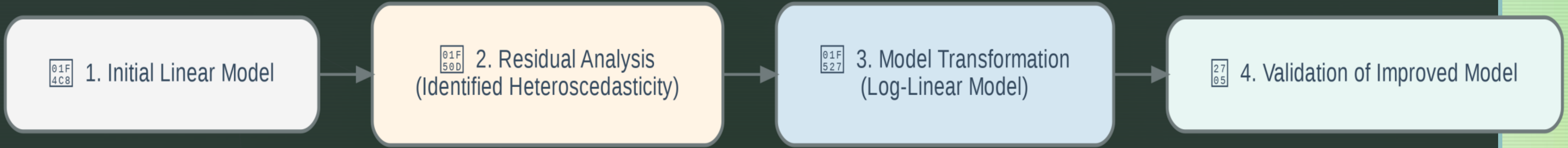
- Our Recommendation : "Focus investment strategy on properties with high rateable values and large outdoor spaces, situated in areas of low unemployment."
- Business Impact
  - ✅ Faster Decision-Making: Created a tool to quickly screen dozens of potential properties using public data.
  - 💰 Reduced Costs: Minimized reliance on expensive initial architectural surveys for every property.
  - 🎯 Risk Mitigation: Built a clear framework for avoiding properties in financially risky or less desirable locations.

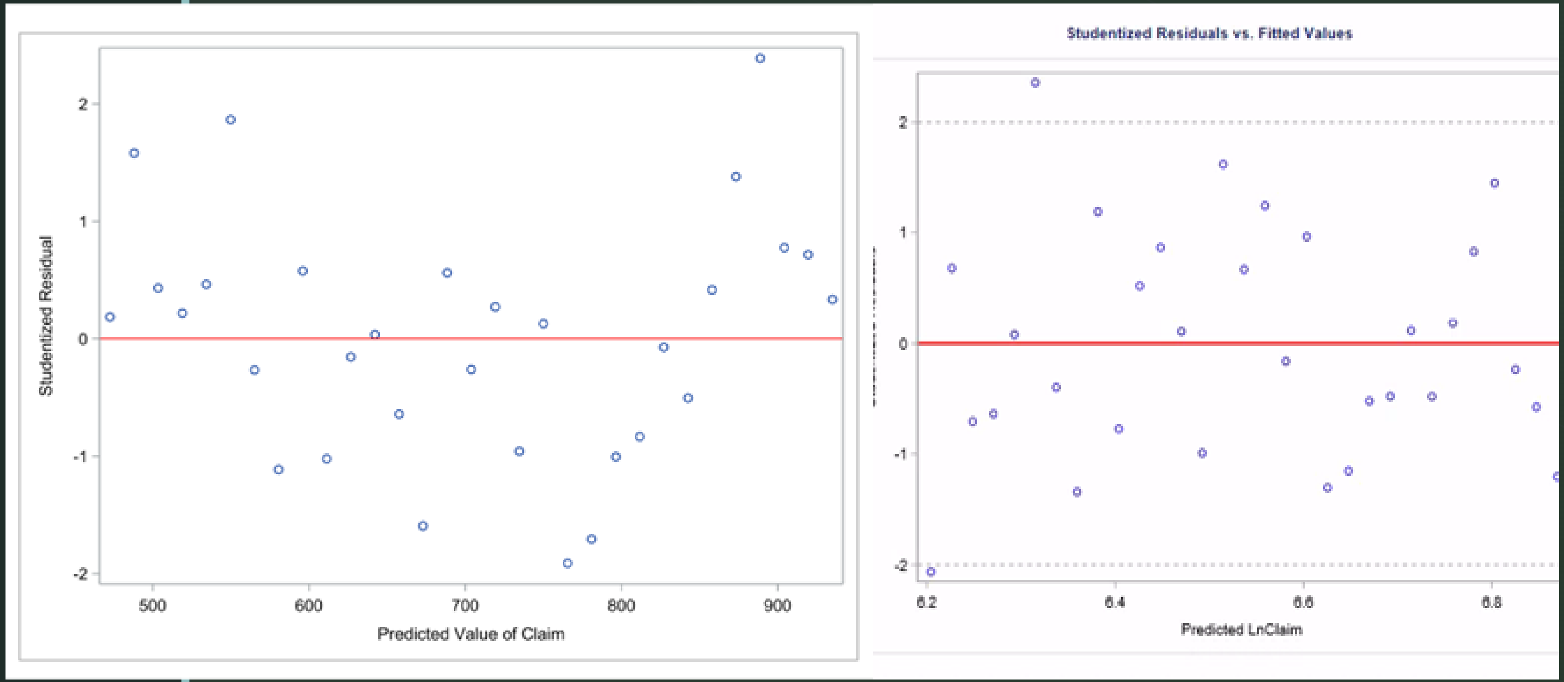
Insurance - Advanced Risk Modeling

The Problem: An insurance company needed to understand how a claimant's age affects their claim amount. An initial linear model was built, but it was flawed and produced unreliable predictions.

The Objective: My task was to diagnose the issues with the initial model and develop a statistically superior one that accurately captures the relationship between age and risk.

My Data-Driven Method: Model Transformation





The original model's 'funnel-shaped' residual plot (left) violated key assumptions. By applying a log transformation, the new model's residuals (right) show a random scatter, confirming its statistical validity and predictive power

Key Finding: Risk Grows Exponentially with Age

📈 Superior Predictive Power

- The improved log-linear model explained 99% of the variance in claim amounts ($R^2 = 0.99$), making it an exceptionally reliable predictive tool.

🔢 Quantifiable Annual Risk

- For every additional year of age, the claim amount increases by approximately 2.2%. This exponential growth was missed by the original model.

Business Impact & Final Recommendation

My Recommendation : "Reject the proposed flat-rate insurance plan. Instead, adopt an age-adjusted pricing strategy based on the log-linear model."

Business Impact

🎯 Accurate Risk Assessment: Prevents the company from under-pricing risk for older, higher-cost customers.	💰 Optimized Pricing Strategy: Created a reliable foundation for developing fair and profitable insurance premiums.	🛡 Reduced Financial Exposure: Mitigated the financial risk associated with mis-pricing an entire customer segment.

The Problem

- A financial services company suspected that younger customers had a higher rate of late payments on personal loans, but they lacked statistical proof to act on this assumption.

The Objective

- My goal was to use hypothesis testing to determine if there was a statistically significant difference in late payment rates between younger (18-30) and older (30+) customers.

Finance - Loan Payment Risk Analysis

Our Data-Driven Method: Chi-Square Test

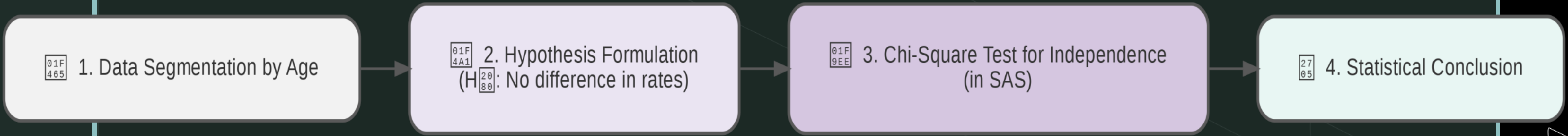
Column 1 Risk Estimates						
	Risk	ASE	95% Confidence Limits		Exact 95% Confidence Limits	
Row 1	0.8233	0.0220	0.7802	0.8665	0.7754	0.8648
Row 2	0.8867	0.0183	0.8508	0.9225	0.8452	0.9202
Total	0.8550	0.0144	0.8268	0.8832	0.8242	0.8822
Difference	-0.0633	0.0286	-0.1195	-0.0072		
Difference is (Row 1 - Row 2)						

Risk Difference Test	
H0: P1 - P2 = 0 Wald Method	
Risk Difference	-0.0633
ASE (H0)	0.0287
Z	-2.2030
One-sided Pr < Z	0.0138
Two-sided Pr > |Z|	0.0276
Column 1 (Late = 0)	

Column 2 Risk Estimates						
	Risk	ASE	95% Confidence Limits		Exact 95% Confidence Limits	
Row 1	0.1767	0.0220	0.1335	0.2198	0.1352	0.2246
Row 2	0.1133	0.0183	0.0775	0.1492	0.0798	0.1548
Total	0.1450	0.0144	0.1168	0.1732	0.1178	0.1758
Difference	0.0633	0.0286	0.0072	0.1195		
Difference is (Row 1 - Row 2)						

Sample Size = 600

The Chi-Square test produced a p-value of 0.0276, which is less than the 0.05 significance level. Therefore, we reject the null hypothesis and conclude there is a statistically significant relationship between age group and late payment status.



📊 A Clear Difference in Rates

The late payment rate for the 18-30 age group was 17.7%.

📉 Compared to Older Customers

The rate for the 30+ age group was significantly lower at 11.3%. This confirmed the company's hypothesis with statistical certainty.

Key Finding: Age is a Significant Predictor of Late Payments

Business Impact & Final Recommendation

Our Recommendation: "**Implement targeted risk mitigation strategies for the 18-30 customer demographic.**"

Business Impact

- 🎯 Proactive Risk Management: Provided the business with a clear, data-backed mandate to address a high-risk segment.
- 💡 Foundation for New Policies: Enabled the development of targeted strategies like stricter payment reminders or adjusted terms for younger applicants.
- 📉 Potential for Reduced Defaults: Created a pathway to lower the overall late payment rate and improve portfolio health.

Core Competencies & Technical Toolkit

Methodologies

- Advanced Regression Modeling
- Hypothesis Testing (T-tests, Chi-Square)
- Statistical Model Diagnostics
- Data Transformation

Tools & Technologies

- SAS
- Data Visualization
- Statistical Analysis

Business Skills

- Data Storytelling
- Business Acumen
- Strategic Recommendations
- Problem Solving

Thank you for reviewing the presentation. I am passionate about using data to solve complex business challenges and am always open to discussing new opportunities.

Syed Taizeem

syedtazeem40@gmail.com

https://viewtaz.netlify.com

Linkedin.com/in/syed-taizeem-hamdani