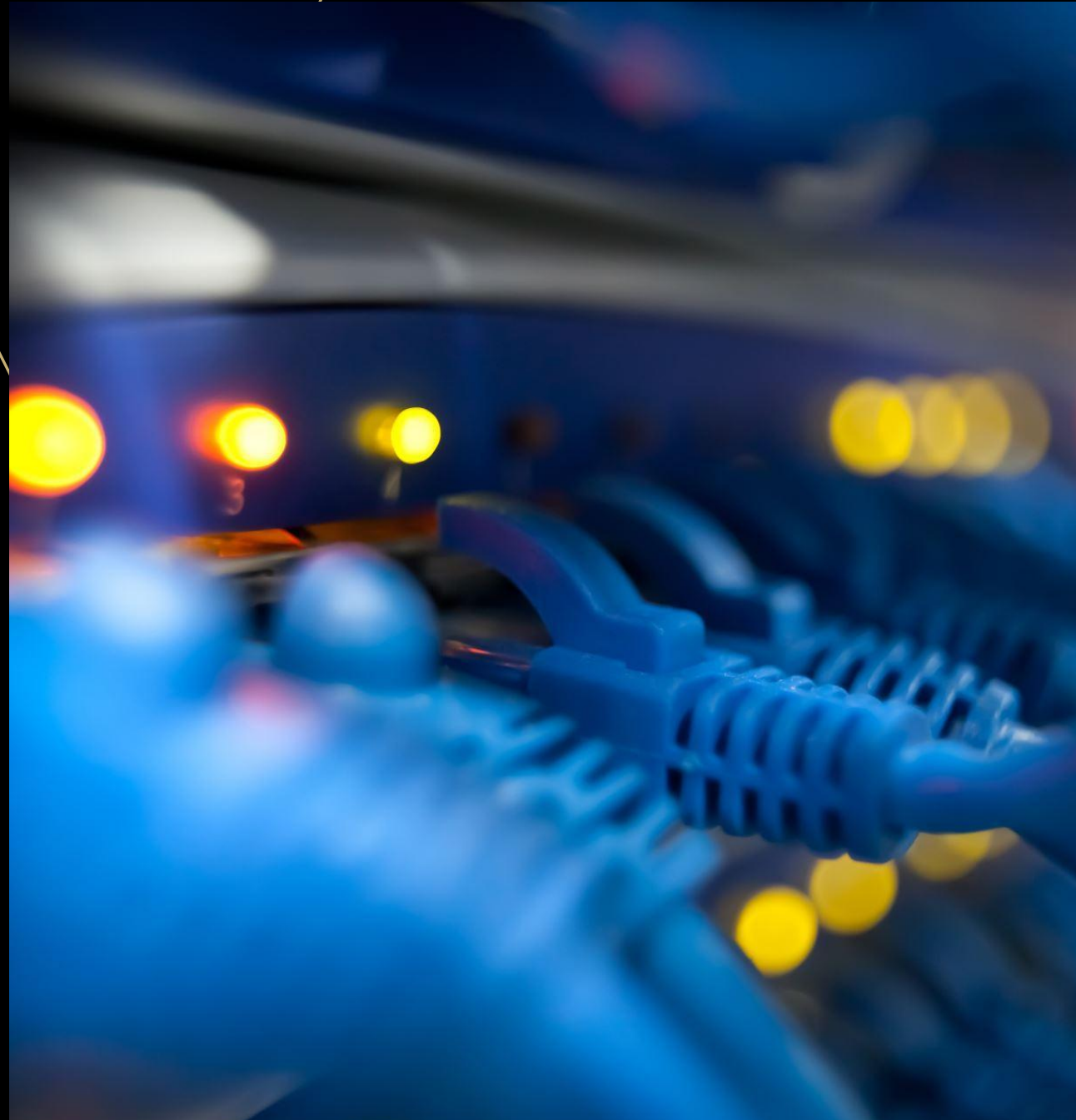
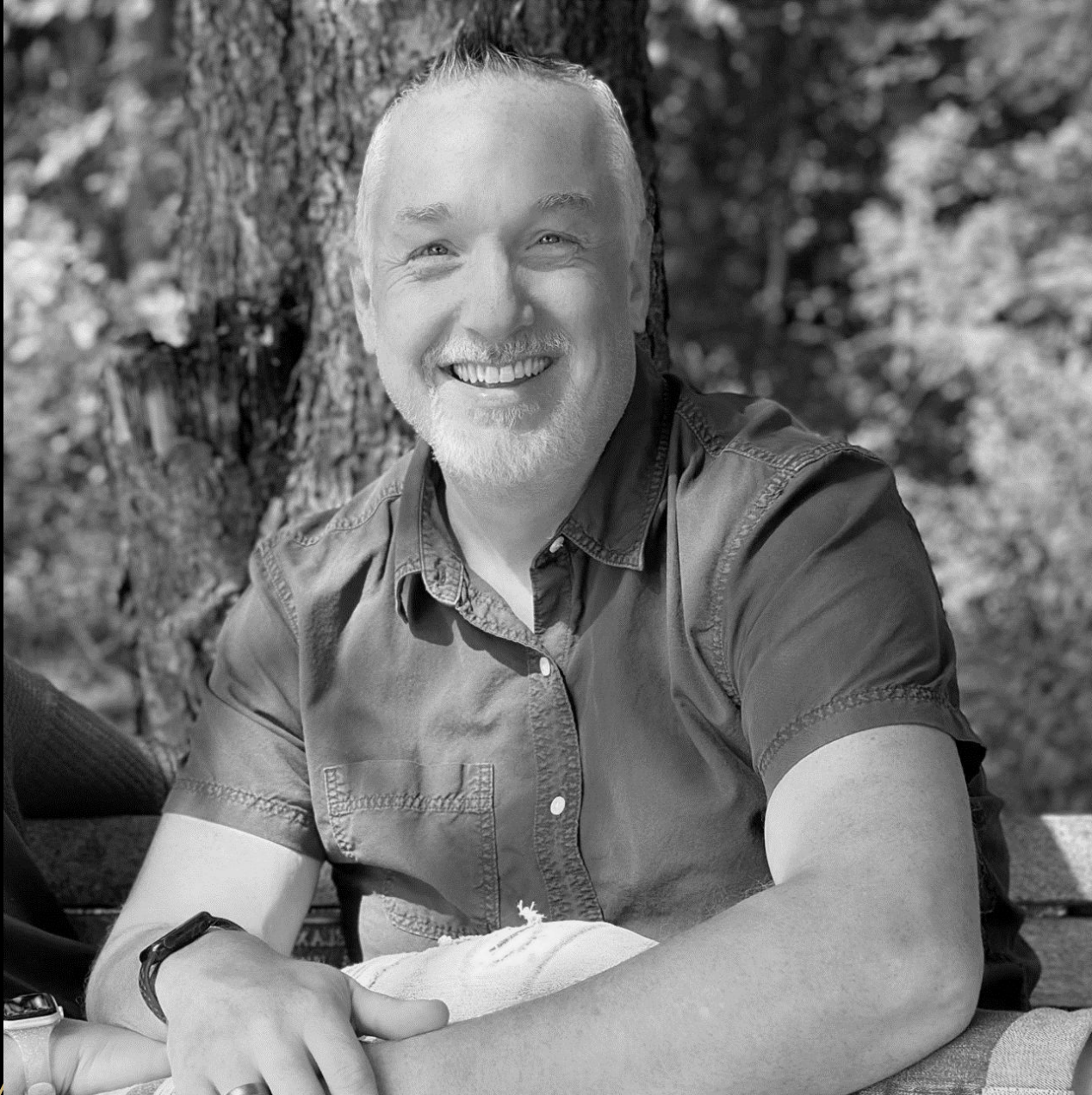


Multiple Linear Regression Analysis on Alarm Cases for Customer Premises Equipment (CPE)

PRESENTED BY KEVIN RUPE

JANUARY 5, 2025





Kevin Rupe

- Bachelor of Science in Mathematics
- Employed at Windstream Communications for 25 years working on many projects dedicated to data analysis and data integrity.
- Currently a Senior Consultant working in the role of Salesforce Admin.
- Seeking Master of Science in Data Analytics

The Problem

- Today's world has more devices online than ever before.
- Everyone wants fast and reliable internet service.
- The companies providing these internet services must act quickly when services are experiencing trouble.
- Not only do customers want their services back online quickly, they want it done right.
- If the Internet Service Provider struggles to keep the customer happy, they will leave and find service elsewhere.



What variables are highly correlated to the alarm being cleared on a case?





Hypothesis

There is a significant correlation between at least one of the independent variables and the alarm being cleared.



Summary of Data Analysis Process

Collect data from
Snowflake
database.

Exclude nulls and
duplicate rows.

Import into
Jupyter Notebook.

Treat outliers using
median imputation.

Convert categorical
variables to
numerical using
encoding.

Remove
multicollinear
variables.

Perform Multiple
Linear Regression
Analysis

Remove
non-statistically
significant variables

Perform final
Multiple Linear
Regression
Analysis



Adjusted R-squared
and R-squared values
are just 4.5%.



Assigned Queue of CSOC
has the strongest negative
effect on the alarm clearing.



Assigned Queue of Outage
has the strongest positive
effect on the alarm clearing.

Outline of Findings

Outline of Findings



High F-statistic with statistically significant p-value assures a good model.



Positive correlations are Primary Customer Condition, Service Status, Enterprise Service Level, Outage Assigned Queue, Open Tier 2 of LAN and Switch, Open Tier 3 of Outage Data, Customer Down, and Customer Degraded.



Negative correlations are Service POD Number, Customer Netcool Origin, Customer Comments, CSOC Assigned Queue, Ethernet Open Tier 2, and Open Tier 3 of Proactive Multiple Underlay, Proactive HA, Proactive 4G, Jitter, Latency, Packet Loss, and SD-WAN HA.



Limitations of Techniques and Tools

- Converting from categorical to numerical adds dimensionality.
- Too many dimensions adds complexity and multicollinearity.
- VIF function doesn't work well with too many variables.
- VIF must be performed numerous times by slowly removing multicollinear variables.



Call to Action

- Reduce complexity, dimensionality, multicollinearity and non-statistically significant variables from the alarm case.
- Create another MLR model evaluating Closing codes instead of Open codes to find a stronger correlation to the alarm clearing.
- Focus on a new dependent variable Service Status rather than Alarm Status with a new MLR model.

What are the expected benefits of the study?

- Assist the ISP in becoming more proactive in its approach to service affecting issues.
- Aid the ISP in streamlining processes, enhancing decision making, and allocating resources.
- All of this provides a positive customer experience while aiming to improve ISP revenue.

