PowerBI Implementation Strategy for Blinkit Customer Churn Analysis

This document details the strategic application of Microsoft Power BI for the exploratory data analysis (EDA) phase of the Blinkit customer churn investigation. Leveraging the `Churn-Dataset.xlsx` dataset, the methodology focuses on establishing a robust data ingestion pipeline, performing conceptual data modeling, and employing standard visualization techniques to uncover initial patterns and correlations associated with customer attrition. This Power BI-driven exploration serves as a complementary visual analysis to the quantitative modeling performed within the associated Jupyter Notebook (`src/blinkit-churn-analysis.ipynb`).

1. Data Ingestion and Model Initialization

1.1. Establishing Data Connectivity

The initial phase involves programmatically connecting Power BI Desktop to the provided Excel data source (`dataset/Churn-Dataset.xlsx`). This is achieved through Power BI's data connector framework, specifically targeting the Excel workbook connector.

- **Process:** The connector parses the specified workbook, identifying available data structures (sheets or named tables). User selection within the Navigator interface designates the target data entity for ingestion.
- **Consideration:** For this analysis, direct loading of the primary data sheet was employed. However, in scenarios involving larger volumes, complex schemas, or known data quality issues, invoking the Power Query Editor (**Transform Data**) prior to loading is the standard best practice. This allows for pre-load data cleansing, shaping, and optimization, reducing the load on the main Power BI data model engine (VertiPaq).

Power BI constructs an in-memory analytical data model based on the ingested data. While the current dataset represents a single flat table, understanding Power BI's modeling capabilities is crucial for scalability and complex analyses.

- **Relational Modeling:** In multi-table scenarios (e.g., separate dimensions for customer demographics, product usage, support interactions), Power BI facilitates the definition of relationships between tables based on common keys. This forms a star or snowflake schema, enabling efficient cross-filtering and analysis across different data facets.
- **Data Type Integrity:** Post-ingestion, verifying and enforcing correct data types (e.g., numerical, text, date) for each column is paramount. This ensures accurate calculations and appropriate behavior within visualizations.
- **DAX for Enhanced Metrics:** The Data Analysis Expressions (DAX) language allows for the creation of sophisticated calculated columns and measures directly within the model. While not strictly necessary for the initial visualizations here, DAX is essential for deriving advanced metrics like dynamic churn rates, customer lifetime value (CLV) predictions (if applicable data exists), or complex segmentations not present in the source data. Examples include creating tenure bins (`SWITCH(TRUE(), [Tenure] <= 6, "0-6 Months", [Tenure] <= 12, "7-12 Months", ...)`) or calculating year-over-year changes.

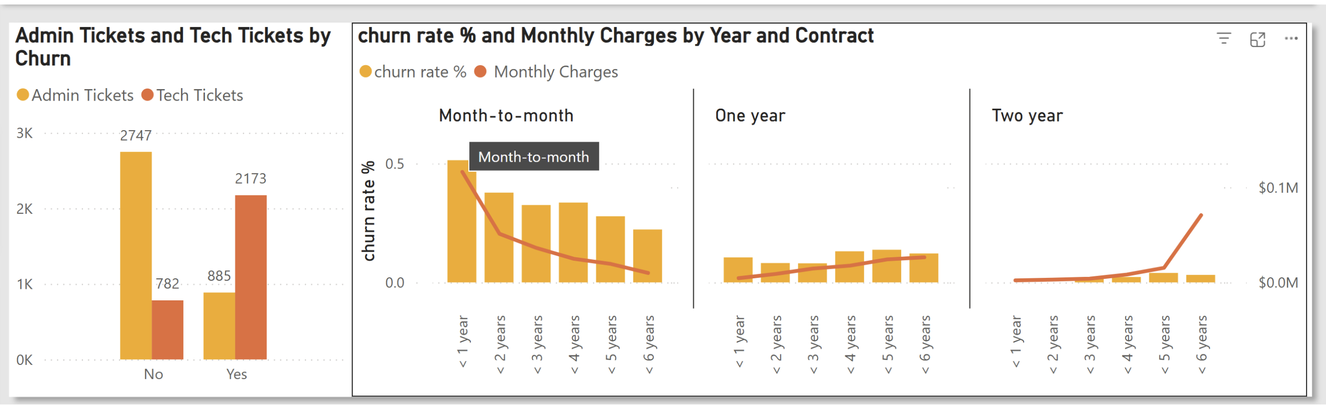
2. Exploratory Visualization Strategy

The primary analytical thrust within Power BI involves the strategic deployment of visualizations to interrogate the data interactively. The goal is to visually identify potential drivers and characteristics of churned customers.

2.1. Analysis of Categorical Predictors

Understanding how churn prevalence varies across different customer segments defined by categorical attributes (e.g., `Gender`, `Contract Type`, `Payment Method`) is a fundamental step.

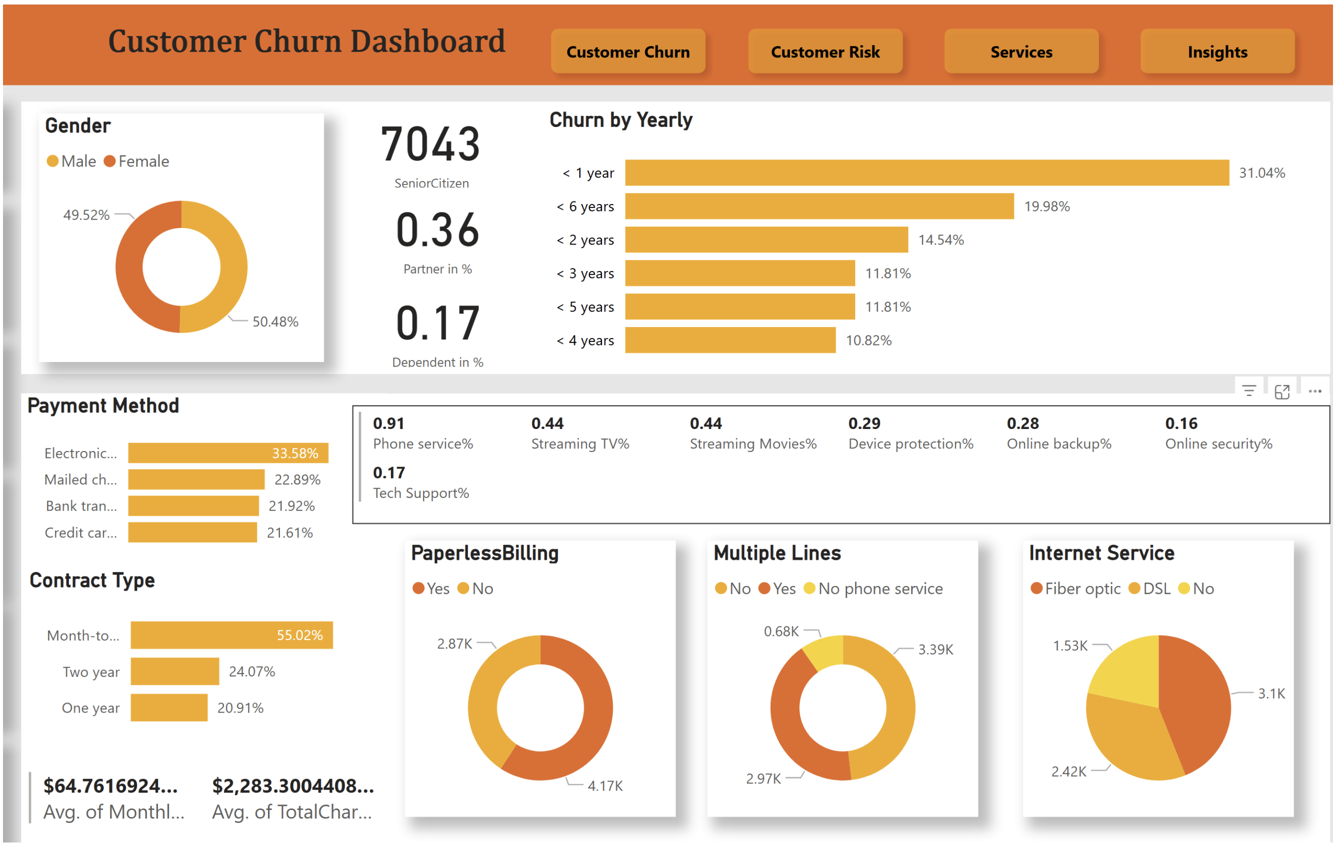
- **Technique:** Employing **Stacked Bar Charts** or **Pie Charts** provides effective visual comparisons of proportions.
- **Configuration Rationale:**
- Categorical Field (`Gender`, `Contract Type`) → Mapped to **Axis** (Bar) or **Legend** (Pie) to define the segments being compared.
- Measure (Count of `UserID` or distinct customers) → Mapped to **Values** to quantify the size of each segment.
- Differentiator (`Churn` Status) → Mapped to **Legend** (Bar) or used to segment the Pie chart, allowing direct visual comparison of churned vs. retained customer counts within each primary category.
- **Interpretation Focus:** Identifying categories with disproportionately high or low churn rates relative to the overall baseline.

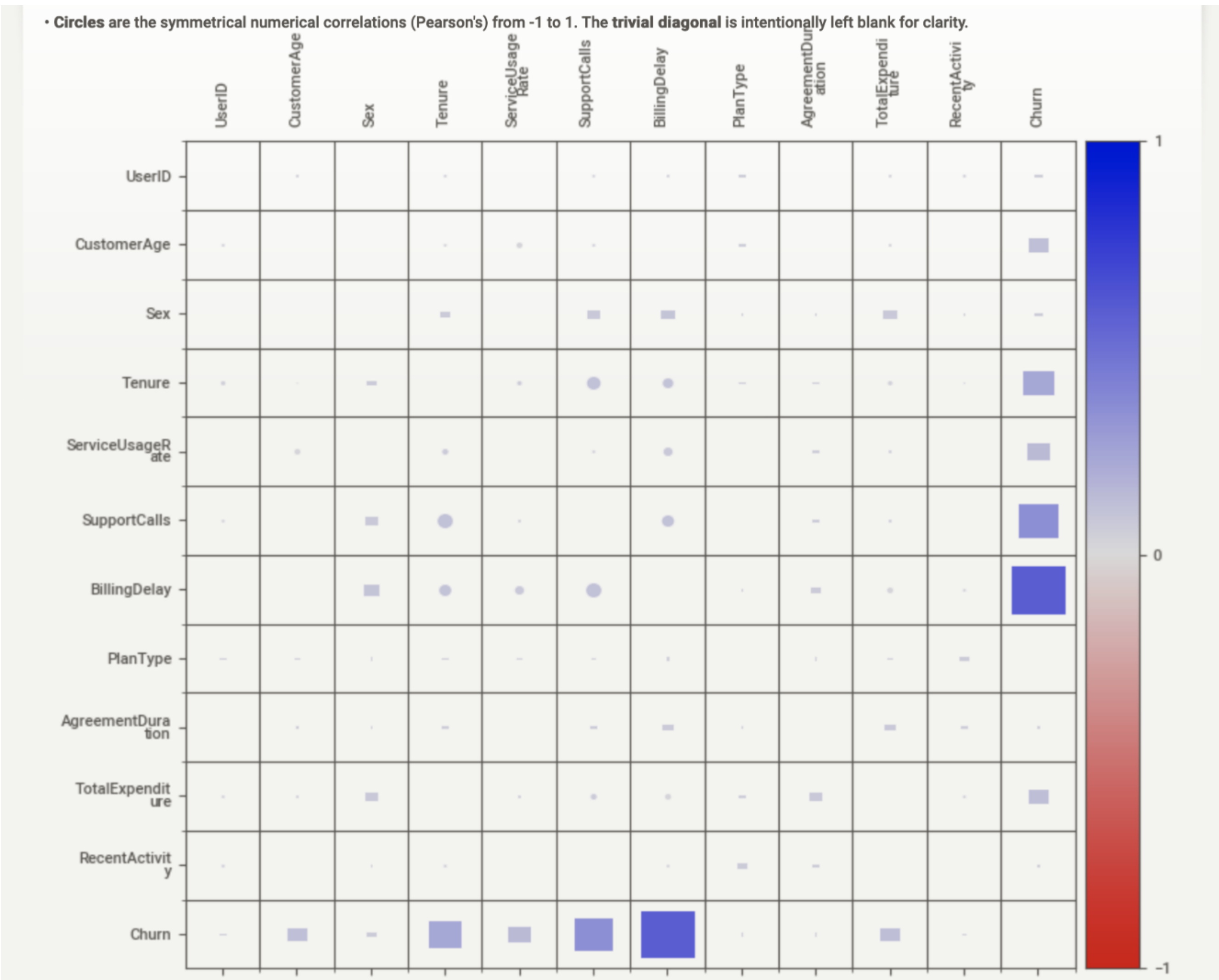


2.2. Analysis of Numerical Predictors

Investigating the influence of continuous numerical variables (e.g., `Tenure`, `MonthlyCharges`, `TotalCharges`) on churn or related metrics requires different visualization approaches.

- **Technique: Line Charts** are suitable for observing trends across ordered numerical categories (like `Tenure`), while **Scatter Charts** excel at revealing correlations between two numerical variables.
- **Configuration Rationale (Line Chart Example - Avg. Monthly Charges vs. Tenure):**
- Ordered Numerical Field (`Tenure`) → Mapped to **Axis** to represent the progression.
- Aggregated Numerical Field (Average of `MonthlyCharges`) → Mapped to **Values** to show the trend of the metric across the axis categories. Aggregation type (Sum, Average, Min, Max) is critical and chosen based on the analytical question.
- Optional Segmentation (`Churn` Status) → Can be mapped to **Legend** to plot separate trend lines for churned and retained customers, highlighting differences in behavior.
- **Interpretation Focus:** Identifying non-linear relationships, inflection points, or distinct patterns associated with churned customers (e.g., do churned customers exhibit significantly higher charges early in their tenure?).

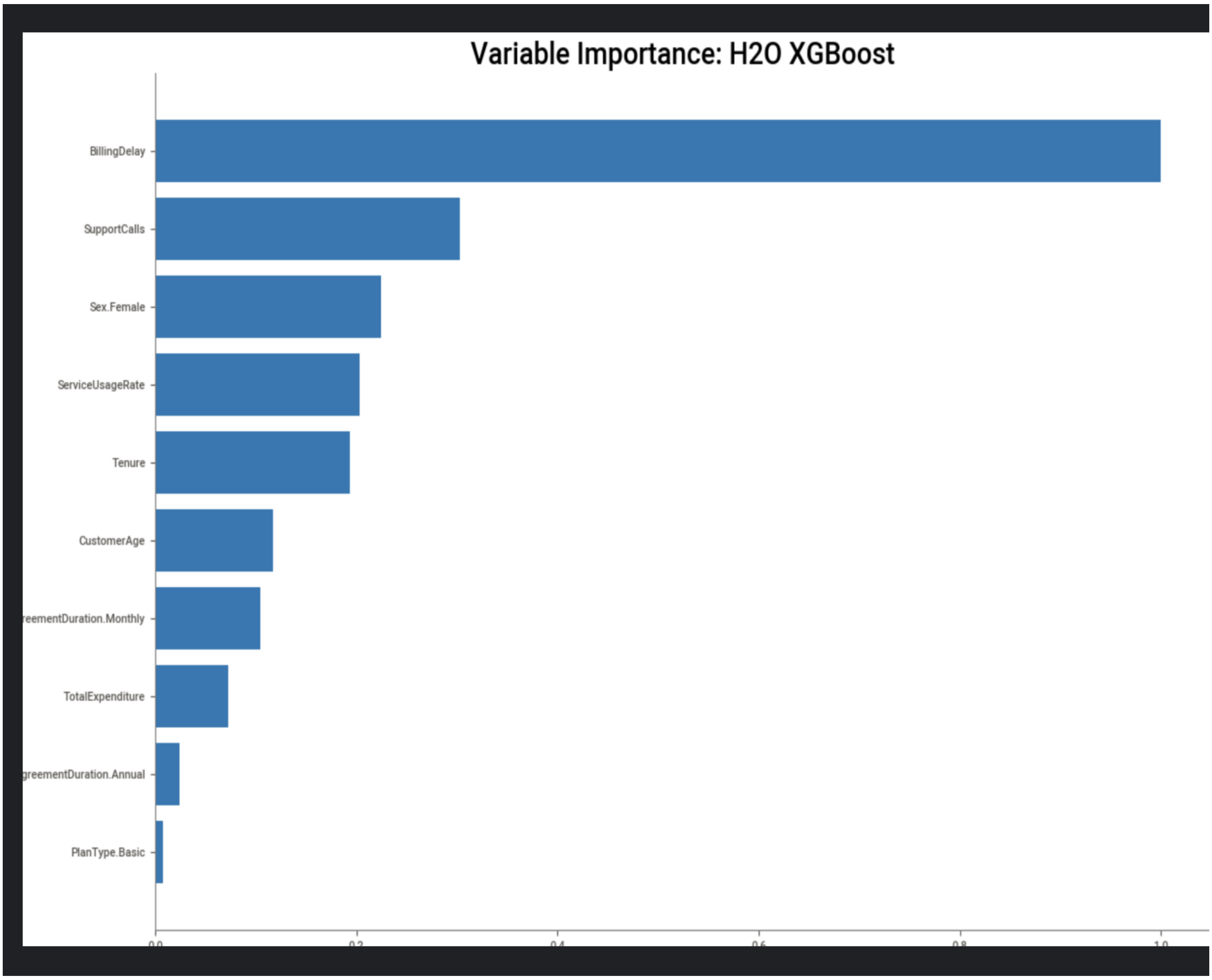

Customer Churn Dashboard
Customer Churn
Customer Risk
Services
Insights
Gender
Male Female
49.52%
50.48%
7043
SeniorCitizen
0.36
Partner in %
0.17
Dependent in %
Churn by Yearly
< 1 year 31.04%
< 6 years 19.98%
< 2 years 14.54%
< 3 years 11.81%
< 5 years 11.81%
< 4 years 10.82%
Payment Method
Electronic... 33.58%
Mailed ch... 22.89%
Bank tran... 21.92%
Credit car... 21.61%
0.91 Phone service%
0.44 Streaming TV%
0.44 Streaming Movies%
0.29 Device protection%
0.28 Online backup%
0.16 Online security%
0.17 Tech Support%
Contract Type
Month-to... 55.02%
Two year 24.07%
One year 20.91%
$64.7616924... Avg. of Monthl...
$2,283.3004408... Avg. of TotalChar...
PaperlessBilling
Yes No
2.87K
4.17K
Multiple Lines
No Yes No phone service
0.68K
3.39K
2.97K
Internet Service
Fiber optic DSL No
1.53K
3.1K
2.42K


• Circles are the symmetrical numerical correlations (Pearson's) from -1 to 1. The trivial diagonal is intentionally left blank for clarity.
UserID
CustomerAge
Sex
Tenure
ServiceUsageRate
SupportCalls
BillingDelay
PlanType
AgreementDuration
TotalExpenditure
RecentActivity
Churn
1
0
-1

2.3. Integrated Dashboard Design Principles

Effective analysis often requires synthesizing insights from multiple visualizations simultaneously.

- **Composition:** Arrange complementary visuals (e.g., overall churn rate KPI, categorical breakdowns, numerical trend lines) onto a single report canvas.
- **Interactivity:** Implement **Slicers** linked to key dimensions (`Service Type`, `Region`, `Demographic Segments`). User interaction with slicers dynamically filters all visuals on the page, enabling rapid exploration of specific sub-populations and their churn characteristics.
- **Cross-Filtering:** Leverage Power BI's default behavior where selecting data points in one visual filters related visuals, facilitating drill-down analysis and discovery of interconnected patterns.



3. Synthesis and Next Steps

The Power BI exploration provides a crucial visual and interactive layer to the churn analysis. It excels at rapidly identifying high-level trends, segment-specific issues, and potential correlations that warrant deeper statistical investigation. Findings from this EDA phase should inform the feature engineering and model selection processes undertaken in the more quantitative analysis environment (e.g., the Python notebook). The interactive nature allows stakeholders to readily explore the data and validate hypotheses generated from other analytical methods.