Customer Engagement Strategies

Subscription Prediction in Banking

Leveraging machine learning to enhance customer targeting and improve marketing ROI

Team Structure and Roles

Overview of key team members

Project Manager/ Scrum Master
Data Engineer
Data Analyst
ML Engineer
Business Analyst

Understanding the significance of optimizing marketing strategies in banking campaigns

In today's highly competitive banking landscape, **inefficient marketing practices** can lead to wasted resources and missed opportunities.

- High marketing spend without adequate returns
- Poor targeting results in low conversion rates
- Customer fatigue from repetitive outreach efforts
- The necessity for more **intelligent** and personalized marketing approaches

To address these issues, it's crucial to leverage data-driven insights and advanced modeling techniques. By identifying and targeting customers who are more likely to respond positively, organizations can **maximize their return on investment (ROI)**. Implementing predictive analytics and uplift modeling will not only enhance targeting precision but also foster customer engagement through meaningful interactions. This approach ensures that marketing budgets are spent effectively, ultimately driving growth and customer satisfaction.

Defining the Core Challenges in Customer Subscription Predictions

The primary challenge in our banking marketing efforts is to **accurately identify** customers who are likely to subscribe after being contacted. This involves several critical components:

- **Targeted Outreach**: It's essential to minimize irrelevant outreach that could lead to customer fatigue, thereby ensuring that only the most promising leads are approached.
- **Predictive Accuracy**: Utilizing advanced modeling techniques will enhance our ability to predict which customers are most likely to engage positively with our offerings.
- **ROI Improvement**: By focusing on high-potential customers, we aim to significantly boost our campaign ROI, ensuring that marketing resources are allocated efficiently and effectively.

Addressing these challenges is vital for transforming marketing strategies and achieving sustainable growth in customer subscriptions.

Integrating Machine Learning and Causal Methods for Effective Subscription Prediction

In this project, we address the challenge of predicting customer subscriptions by treating it as a churn analogue. Our approach combines machine learning with causal inference to create a robust framework that delivers interpretable and actionable insights. Key elements of our methodology include:

- **Predictive Modeling:** Utilizing various algorithms to identify customers who are more likely to subscribe based on their interaction history.
- **Causal Analysis:** Implementing uplift modeling to assess the incremental impact of marketing outreach on subscription rates.
- **Deployment-Ready Solutions:** Developing a user-friendly interface for stakeholders to access insights, ensuring compliance and protecting customer privacy.

By integrating these methodologies, we aim to enhance targeting strategies, reduce marketing inefficiencies, and ultimately drive better ROI for banking campaigns. This project serves as a foundation for future advancements in customer relationship management and ethical AI practices.

Key objectives of our predictive modeling initiative are outlined below.

The primary goals of our project are:

- **Build Predictive Models:** Develop robust models to accurately predict customer subscription behavior.
- **Perform Uplift/Survival Analysis:** Understand the incremental impact of marketing efforts on customer behavior.
- **Create ROI Dashboard:** Design a comprehensive dashboard to visualize return on investment and campaign effectiveness.
- **Provide Business Recommendations:** Generate actionable insights and recommendations to enhance targeting strategies.

Our approach emphasizes the integration of interpretable machine learning methods with causal inference, ensuring that the resulting models are not only effective but also understandable to stakeholders. By utilizing these objectives, we aim to drive smarter marketing strategies, ultimately leading to improved customer engagement and higher campaign ROI.

Data Sources Overview for Prediction Models

Key datasets for effective analysis

UCI Bank Marketing Dataset

The UCI Bank Marketing dataset comprises **45,000 rows** of customer data. This dataset includes various features related to customer demographics, marketing campaigns, and subscription outcomes, enabling us to analyze marketing effectiveness and predict customer behavior accurately.

ECB Macro Indicators

The European Central Bank macro indicators provide essential economic context, including interest rates, inflation rates, and unemployment figures. By incorporating these indicators, we can better understand how macroeconomic conditions influence customer behavior and marketing campaign success.

Temporal Joining for Context

Temporal joining of datasets allows us to connect the UCI Bank Marketing data with macroeconomic indicators over time. This contextual integration helps identify trends and correlations, enhancing the predictive power of our models and ensuring robust analysis.

Data Preparation Process for Predictive Modeling

Ensuring Clean, Safe, and Effective Datasets

Privacy Preservation Measures

To protect sensitive information, we remove any personally identifiable information (PII) from the dataset. This step ensures compliance with data privacy regulations and safeguards customer trust, ultimately enabling us to focus on relevant predictive features without compromising individual privacy.

Handling Missing Data Effectively

Missing values can significantly impact model performance. We utilize imputation techniques to fill gaps in the data, ensuring continuity and completeness. This process helps maintain data integrity, allowing for more accurate predictions and improved model reliability during analysis and deployment.

Categorical Data Preparation

Categorical variables require encoding to facilitate analysis by machine learning models. We employ methods such as one-hot encoding and label encoding to transform these variables into numerical formats. This step is essential for creating a dataset that is suitable for algorithmic processing and modeling.

Methodology Overview: A Comprehensive Workflow for Effective Subscription Prediction

The methodology for our project is structured around a systematic workflow designed to maximize predictive accuracy and business impact. The key steps include:

- **ETL (Extract, Transform, Load):** Gather and prepare data from multiple sources.
- **EDA (Exploratory Data Analysis):** Analyze datasets to understand patterns and relationships.
- **Feature Engineering:** Create relevant features that enhance model performance.
- **ML (Machine Learning):** Deploy various algorithms for predictive modeling.
- **Uplift Modeling:** Determine the incremental impact of marketing interventions on customer behavior.
- **Survival Analysis:** Assess customer retention and subscription longevity.
- **Dashboard Development:** Build a user-friendly dashboard to visualize insights and metrics.
- **Deployment:** Implement the solution in a production environment for real-time usage.

This comprehensive approach ensures that each phase is meticulously executed, leading to a robust, interpretable model pipeline that drives effective marketing strategies.

Modeling Techniques for Subscription Prediction

Exploring algorithms for effective insights

Logistic Regression: An Interpretative Approach

Logistic Regression serves as a foundational tool due to its **simplicity and clarity**. It allows interpretable results, making it easier to understand the relationship between features and subscription likelihood, which is crucial for strategic decision-making in marketing campaigns.

Advanced Techniques: Random Forest and Boosting

Random Forest, LightGBM, and XGBoost provide **high performance** for complex datasets. These ensemble methods enhance predictive accuracy through multiple decision trees, effectively managing nonlinear relationships while minimizing overfitting, thus driving better results in customer targeting.

Causal Inference: T-/X-Learner and Causal Forest

The T-/X-Learner and Causal Forest models facilitate **causal analysis** by estimating treatment effects. They help isolate the impact of marketing interventions, enabling a clearer understanding of customer behavior and offering actionable insights to optimize outreach strategies.

🏛️ Bank Marketing Campaign Analysis Dashboard

📈 Model Comparison

Model Performance Comparison

Feature Engineering: Key Creation Techniques

Enhancing Predictive Accuracy for Targeting

Recency-Frequency Metrics

Recency-frequency metrics are essential for identifying customer engagement levels. By analyzing how recently and frequently customers have interacted with campaigns, we can derive insights into their likelihood of responding positively to future outreach, ultimately enabling us to target the right individuals more effectively.

Campaign Count Analysis

Campaign count analysis tracks how many marketing campaigns each customer has been exposed to within a specified timeframe. This data helps assess customer fatigue and response patterns, allowing us to fine-tune our strategies and optimize outreach to maximize engagement and reduce wasted marketing spend.

Macroeconomic Lags

Incorporating macroeconomic lags involves analyzing external economic factors such as interest rates and inflation over time. Understanding these influences can enhance subscription predictions by providing context to customer behavior changes, thereby allowing us to adjust marketing strategies in response to evolving economic conditions.

Validation and Evaluation Metrics Overview

Assessing the effectiveness of our models

Time-Aware Cross-Validation

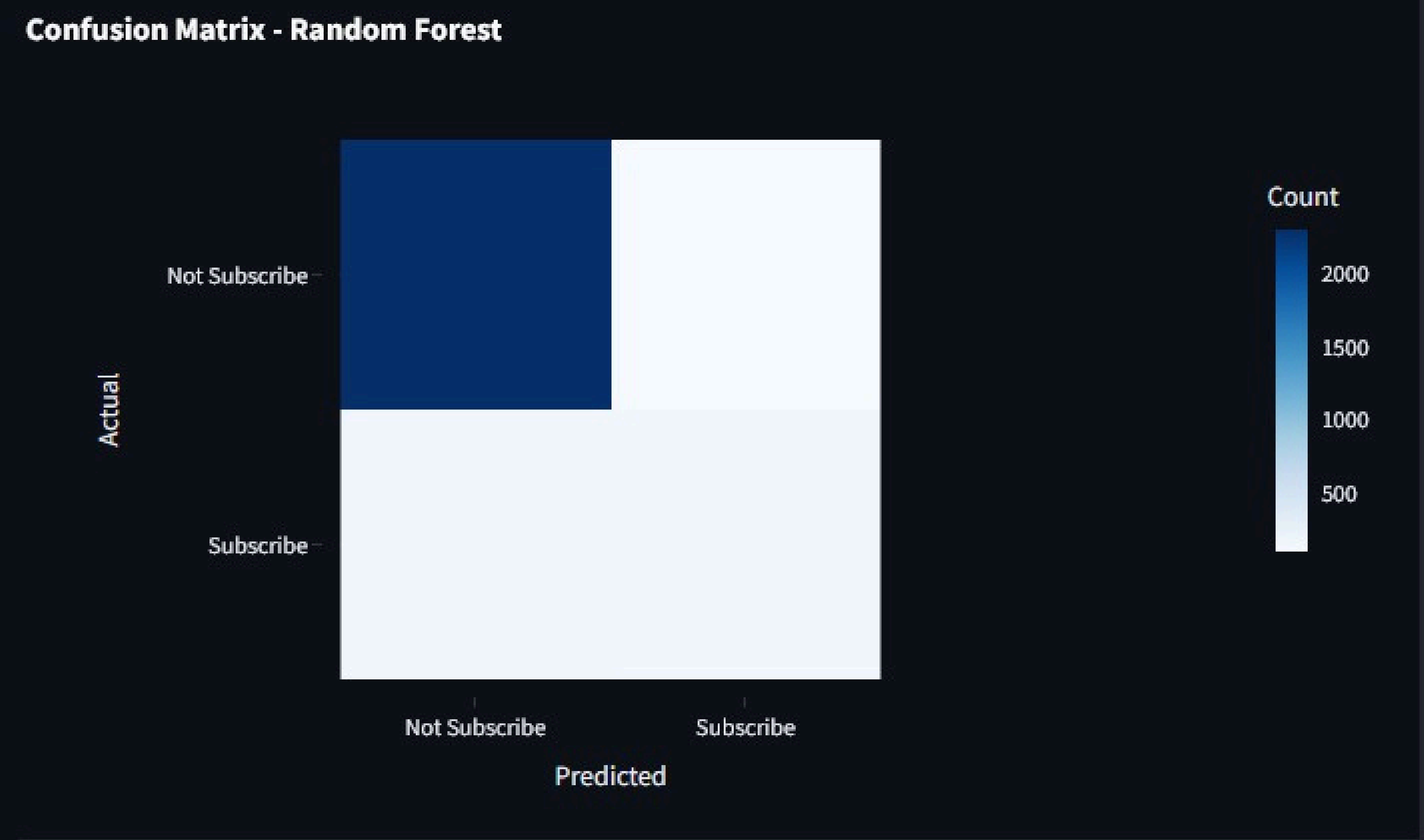
Implementing time-aware cross-validation ensures that our models are tested against **chronologically ordered data**, preserving the temporal nature of banking campaigns. This approach mitigates leakage of future information into the training set, providing a realistic assessment of model performance in real-world scenarios.

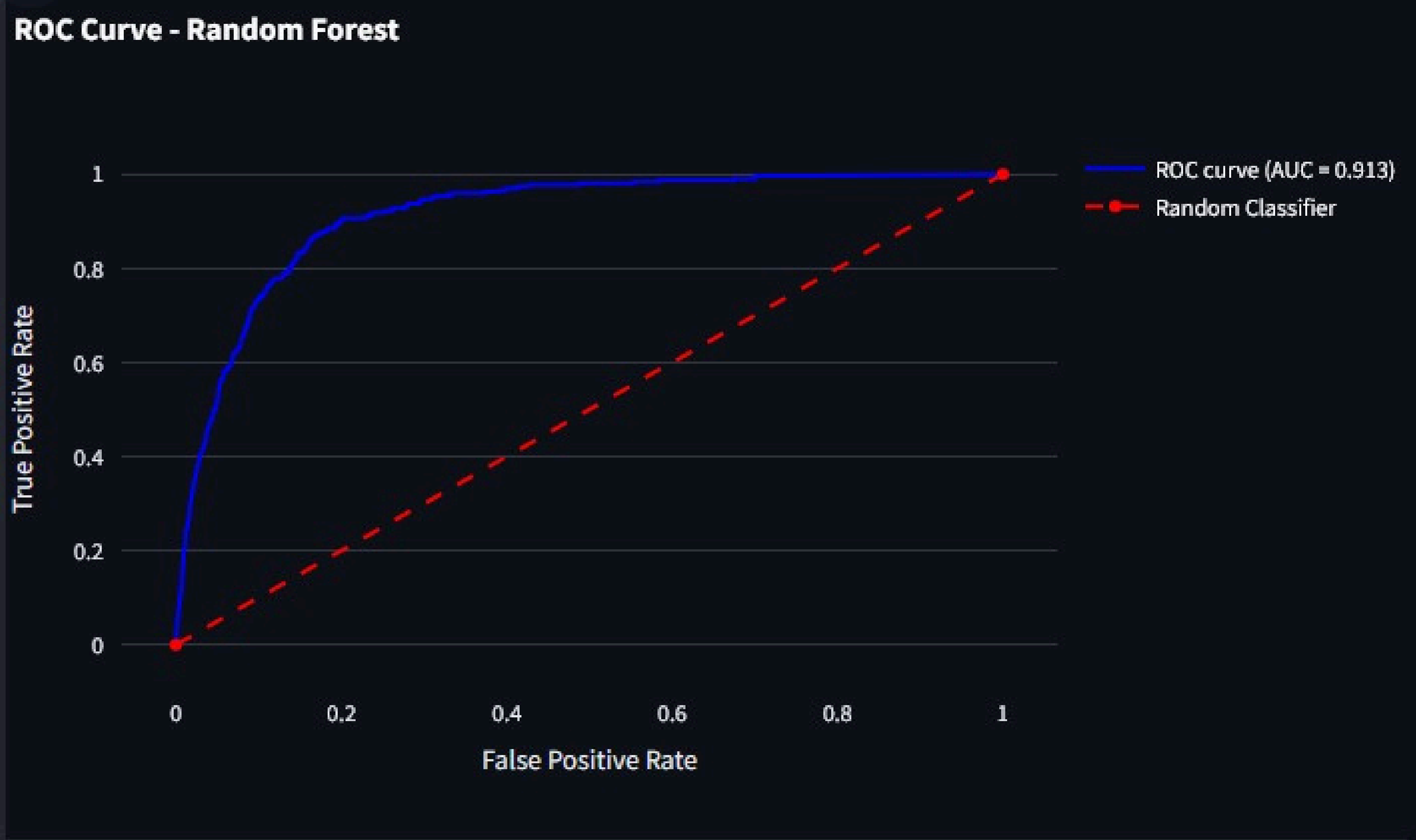
Performance Metrics

To evaluate our models, we utilize key performance metrics such as **PR-AUC and ROC-AUC**, which offer insights into the models' ability to distinguish between subscribing and non-subscribing customers. Additionally, Precision@K assesses the effectiveness of targeting within the top K predictions, enhancing marketing strategy precision.

Uplift and Incremental Analysis

We conduct uplift Qini and AUUC analyses to determine the added value of our marketing interventions. By evaluating incremental lift analysis, we can isolate the specific impact of our campaigns, allowing for **data-driven adjustments** to optimize future marketing efforts and improve overall ROI.





Classification Report

	precision	recall	f1-score	support
0	0.9317	0.9582	0.9448	2,393
1	0.6032	0.475	0.5315	320
accuracy	0.9012	0.9012	0.9012	0.9012
macro avg	0.7675	0.7166	0.7381	2,713
weighted avg	0.893	0.9012	0.896	2,713

Detailed Performance Metrics

	Model	Accuracy	ROC-AUC	Precision	Recall	F1-Score
0	Logistic Regression	0.842	0.903	0.413	0.819	0.549
1	Random Forest	0.901	0.913	0.603	0.475	0.531
2	Gradient Boosting	0.889	0.912	0.522	0.672	0.587
3	SVM	0.215	0.510	0.113	0.822	0.198

🎯 **Best Performing Model: Random Forest**

Accuracy
0.901

ROC-AUC
0.913

F1-Score
0.531

Visualization and Dashboard Insights

Analyzing data for strategic decision-making

Power BI Dashboard Overview

The Power BI dashboard presents a comprehensive view of customer engagement metrics, enabling stakeholders to visualize key performance indicators. This interactive tool facilitates informed decision-making by allowing users to track campaign performance and make necessary adjustments in real-time based on insights gathered from the data.

Uplift Deciles Analysis

Uplift deciles allow us to segment customers based on their predicted response to marketing campaigns. By understanding the varying levels of responsiveness, we can allocate resources more effectively, maximizing marketing impact while minimizing wasteful spending on less likely subscribers.

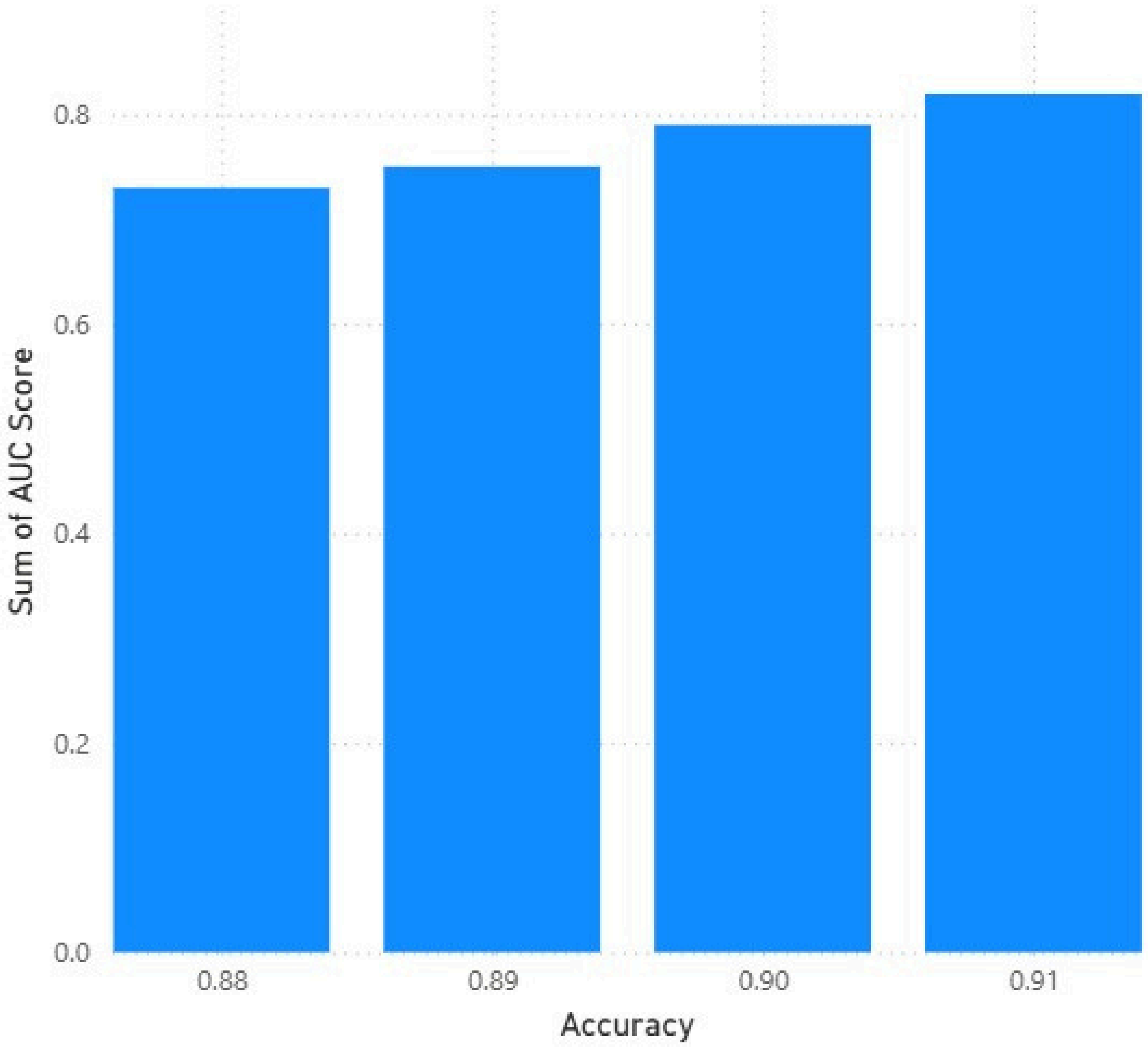
ROI Simulator Functionality

The ROI simulator empowers teams to project the financial outcomes of different marketing strategies. By simulating various scenarios, users can assess potential return on investment, adjust their tactics accordingly, and ensure resources are directed towards the most promising initiatives.

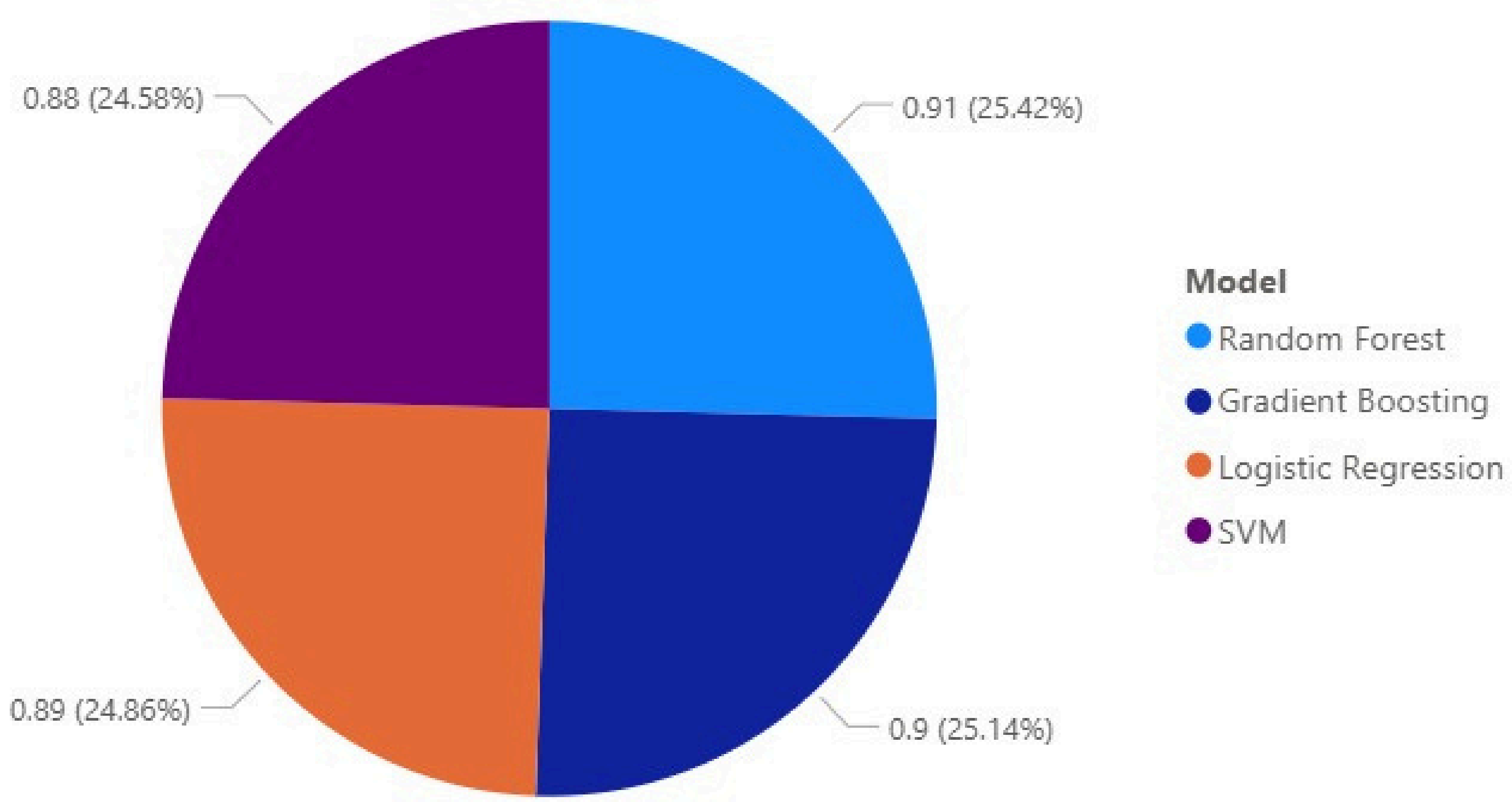
Model	Accuracy	AUC Score
SVM	0.88	0.73
Logistic Regression	0.89	0.75
Gradient Boosting	0.90	0.79
Random Forest	0.91	0.82

Sum of Accuracy by Model

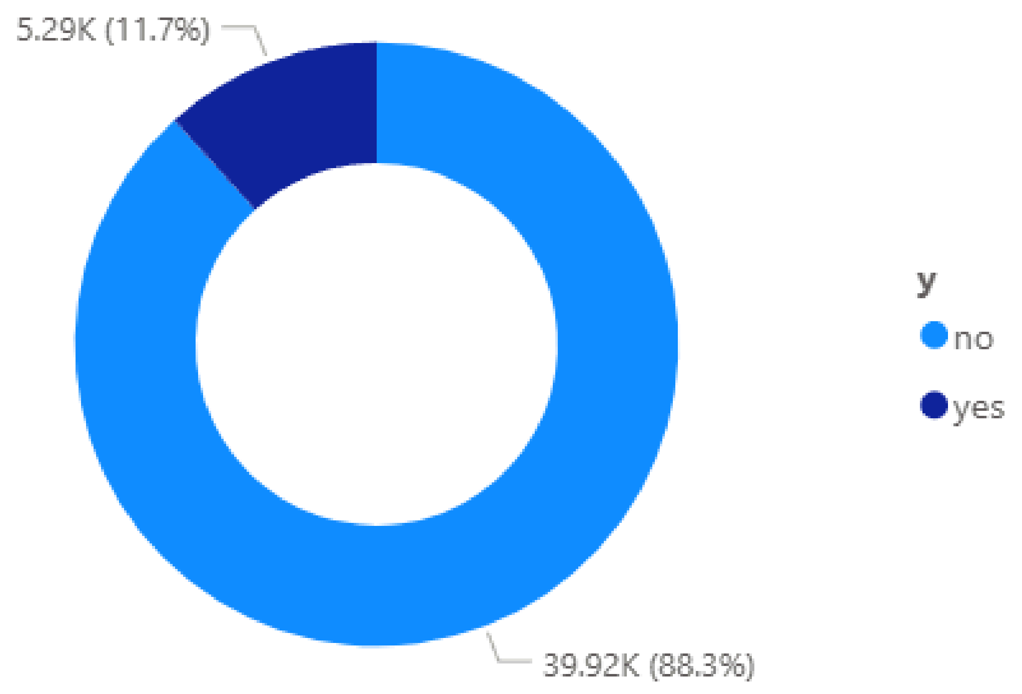
Sum of AUC Score by Accuracy



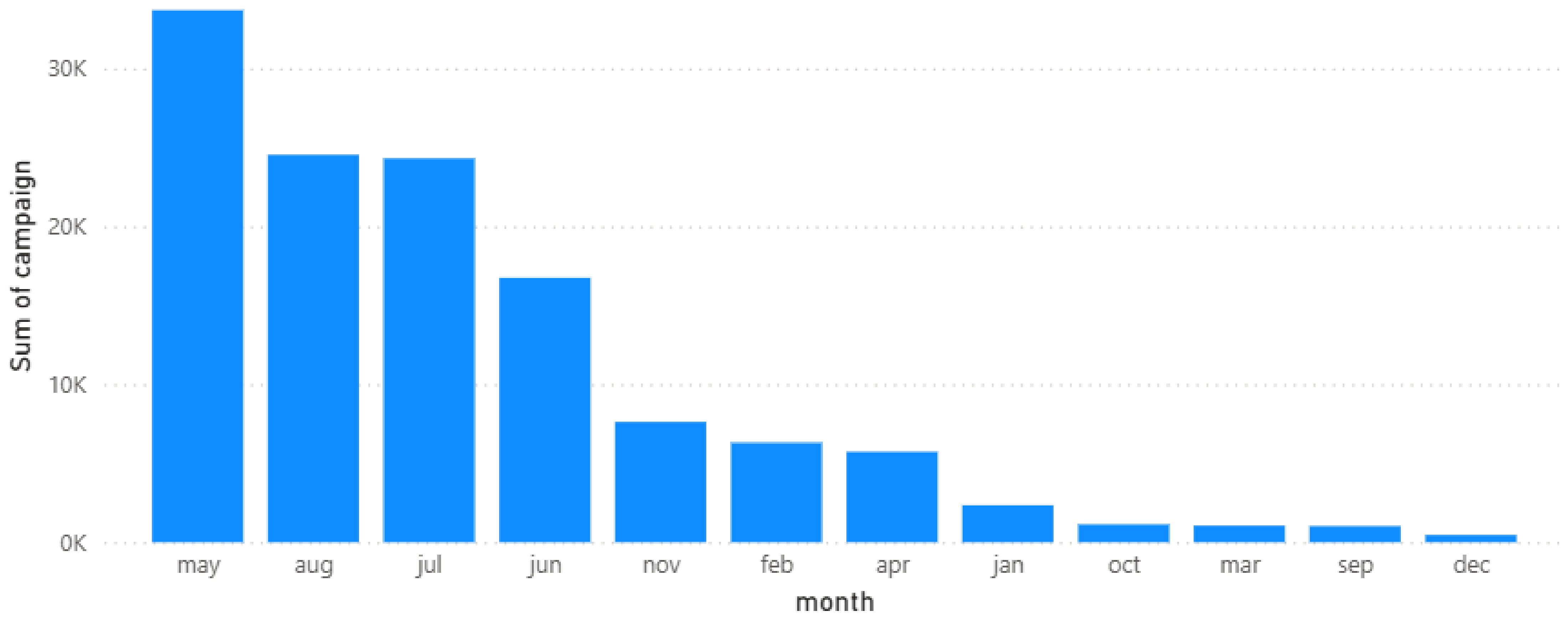
Sum of Accuracy by Model



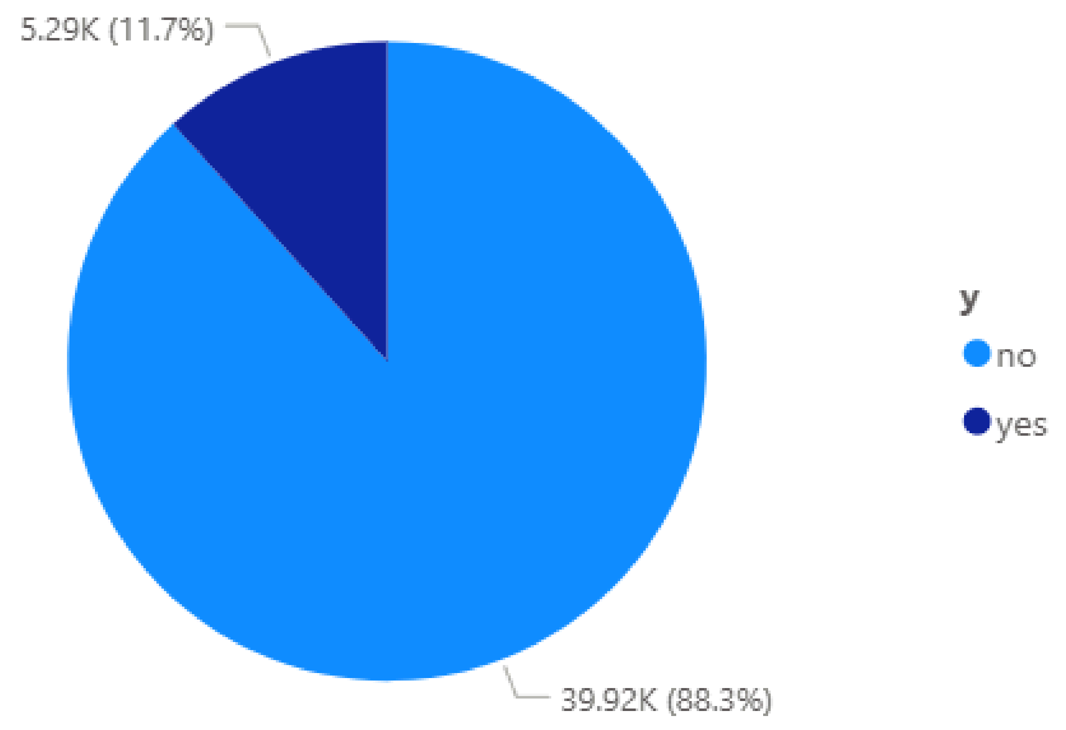
Count of y by y



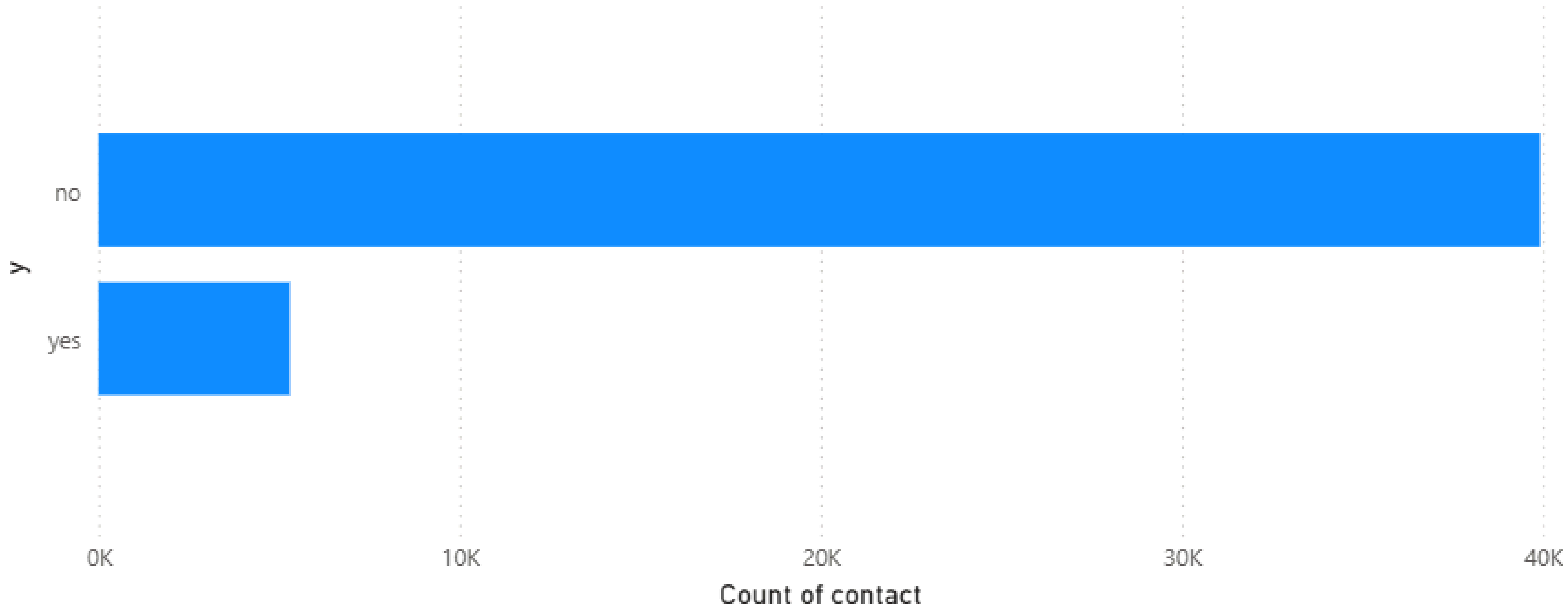
Sum of campaign by month



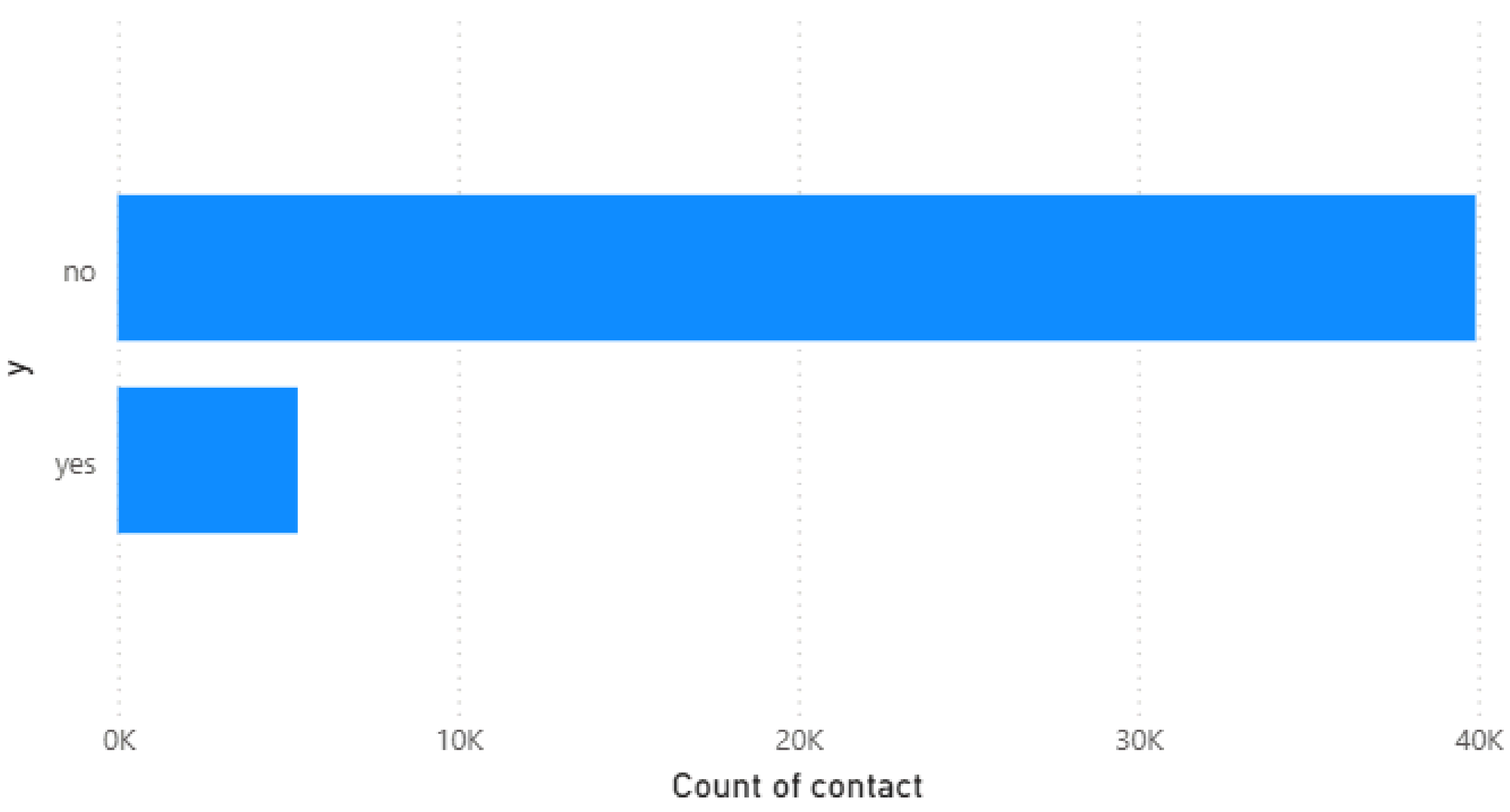
Count of y by y



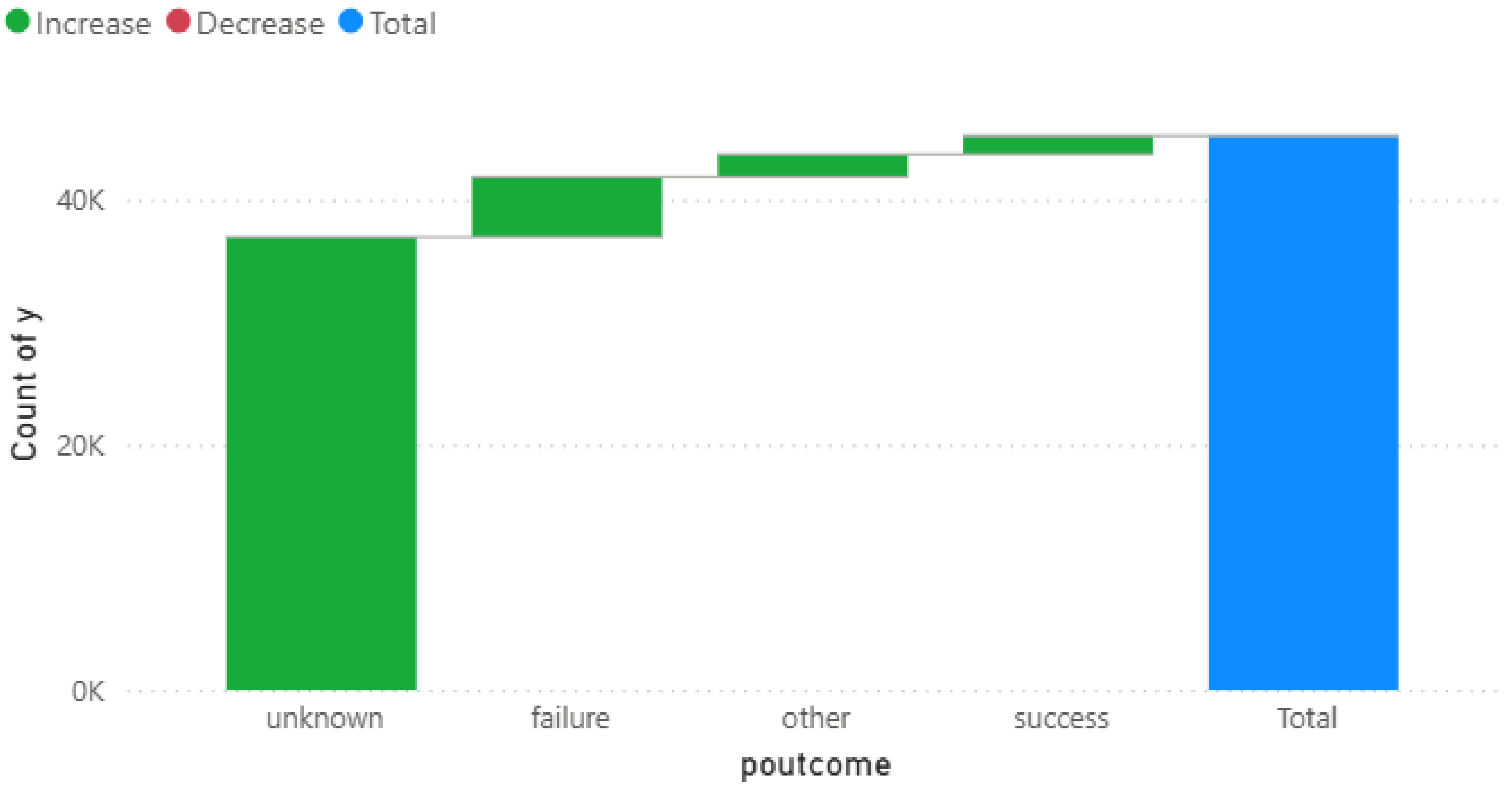
Count of contact by y



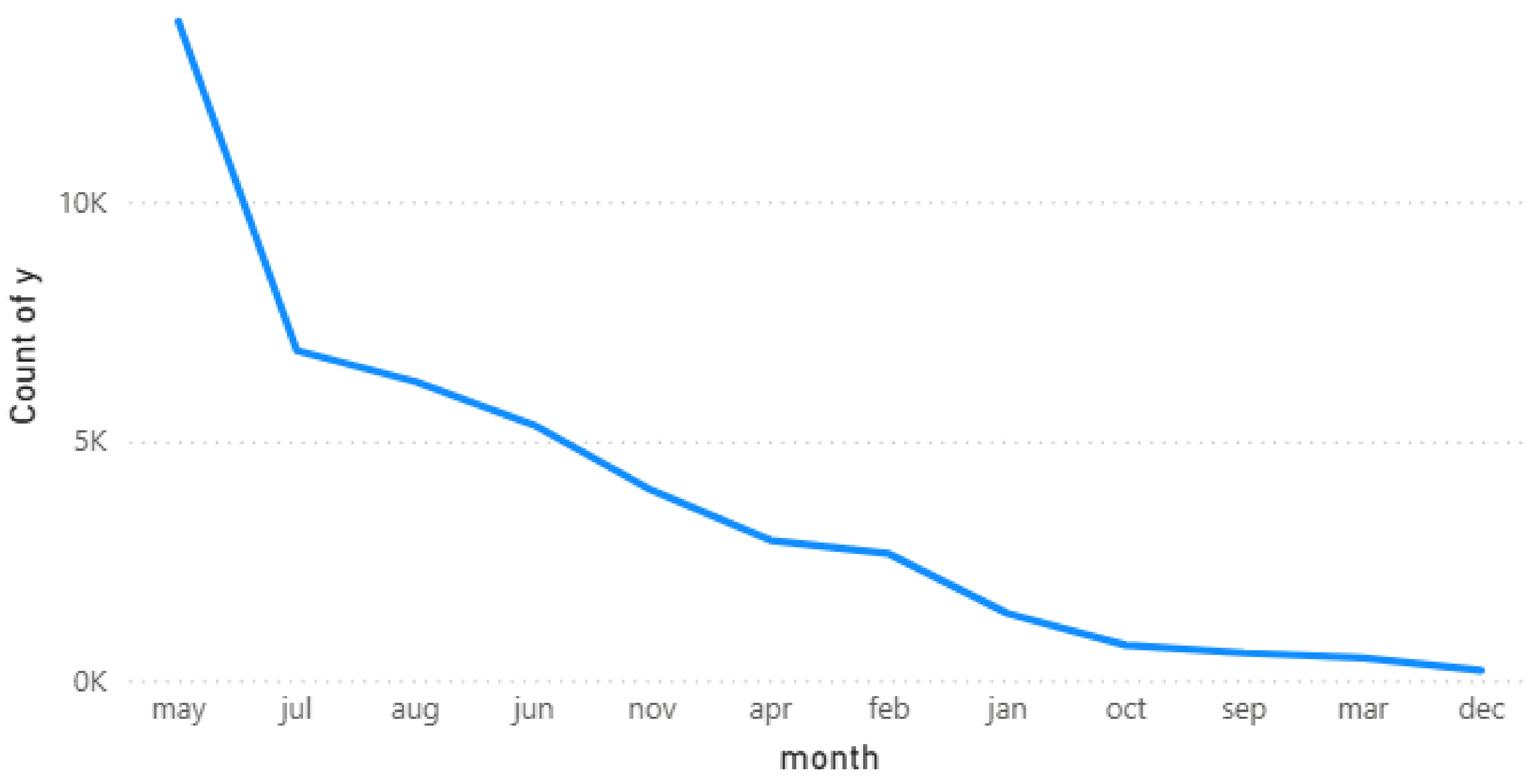
Count of contact by y



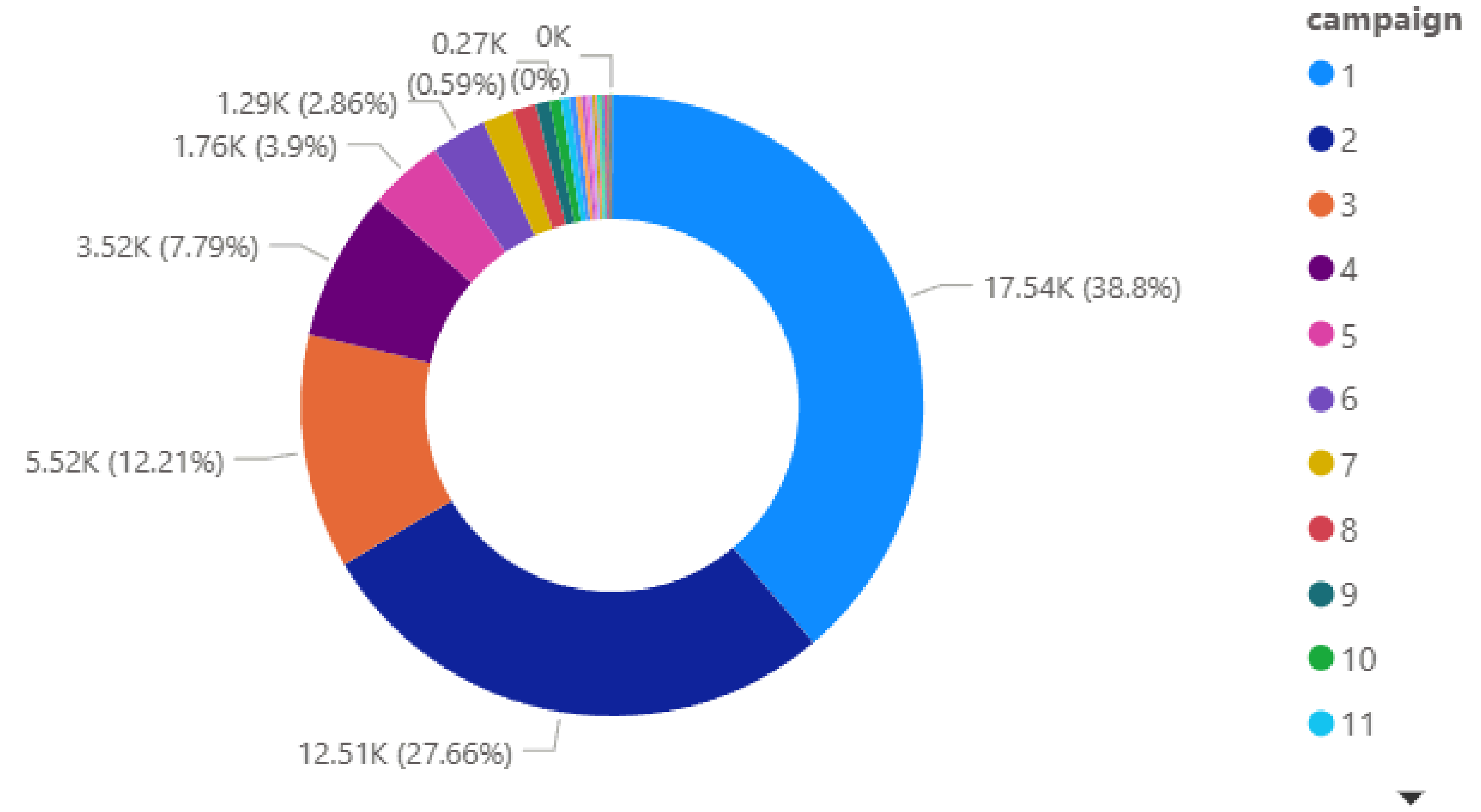
Count of y by poutcome



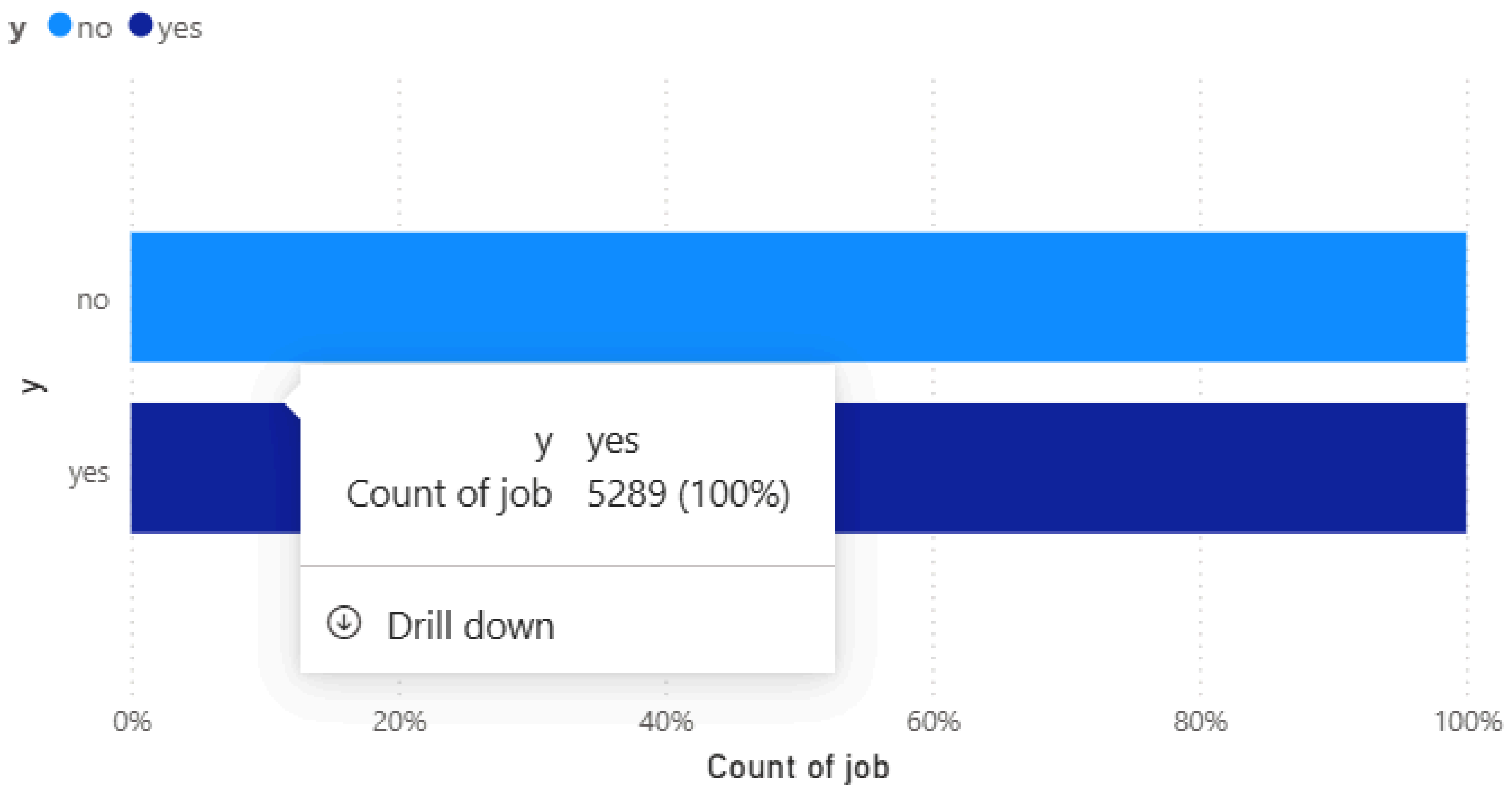
Count of y by month



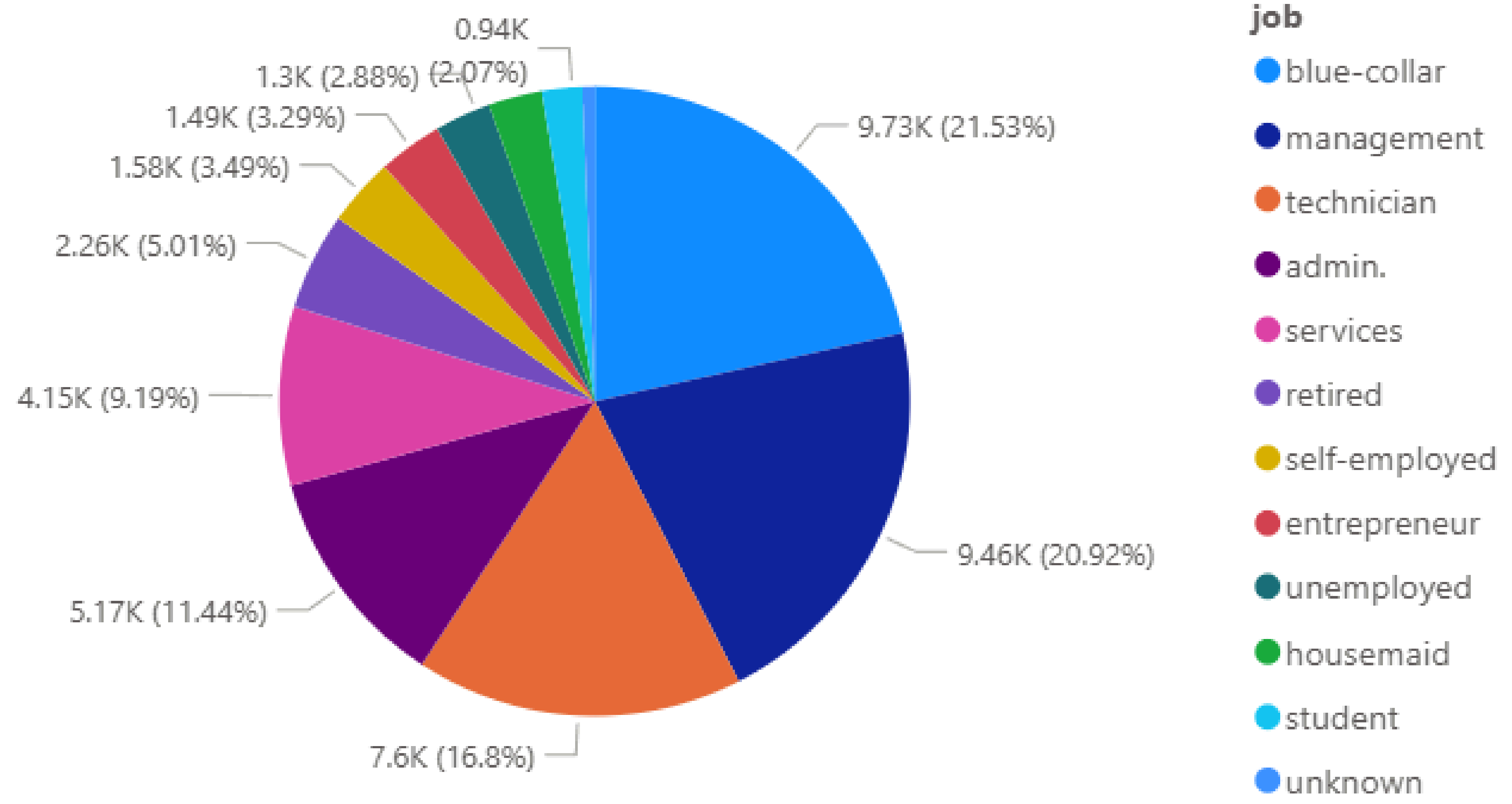
Count of y by campaign



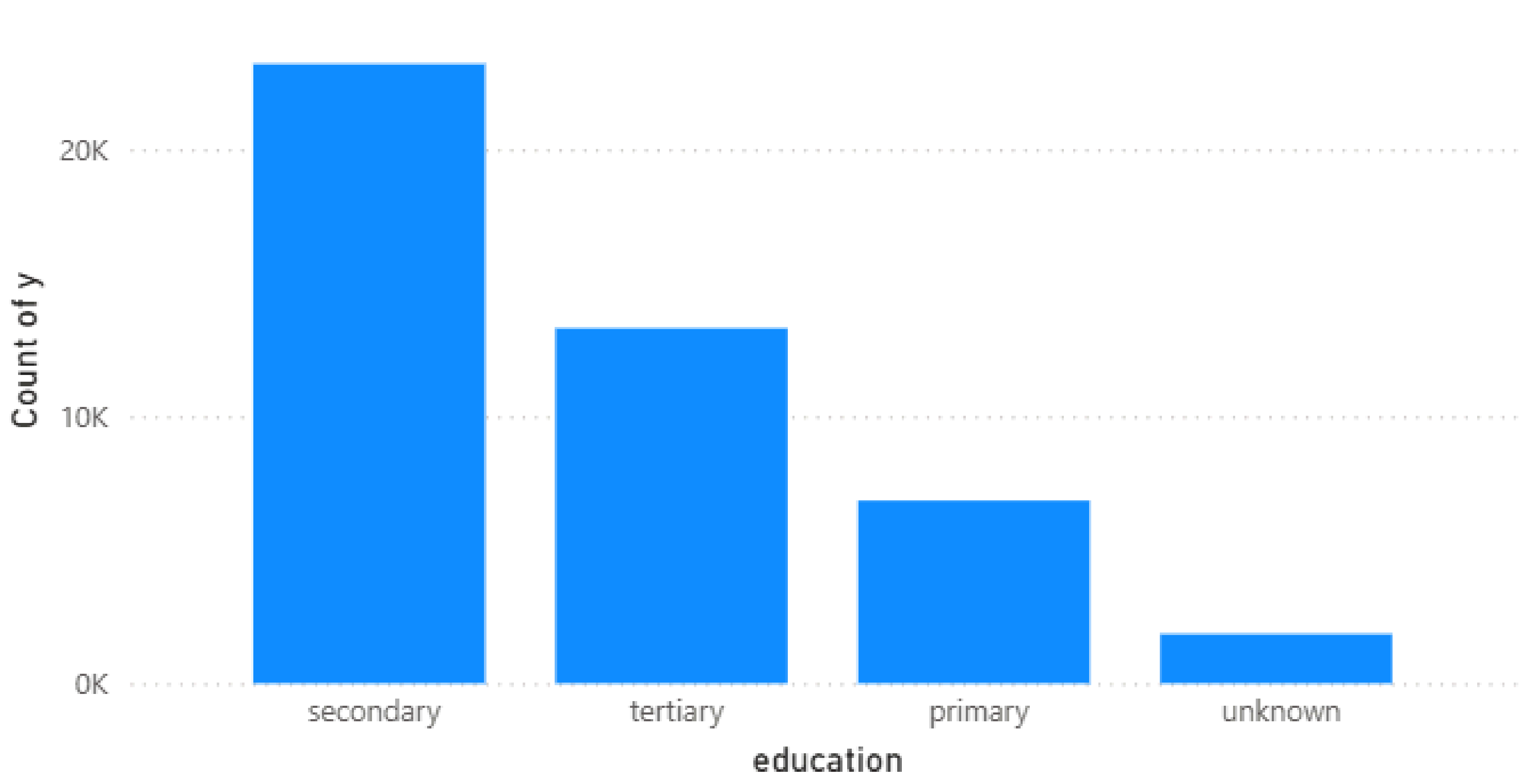
Count of job by y and y



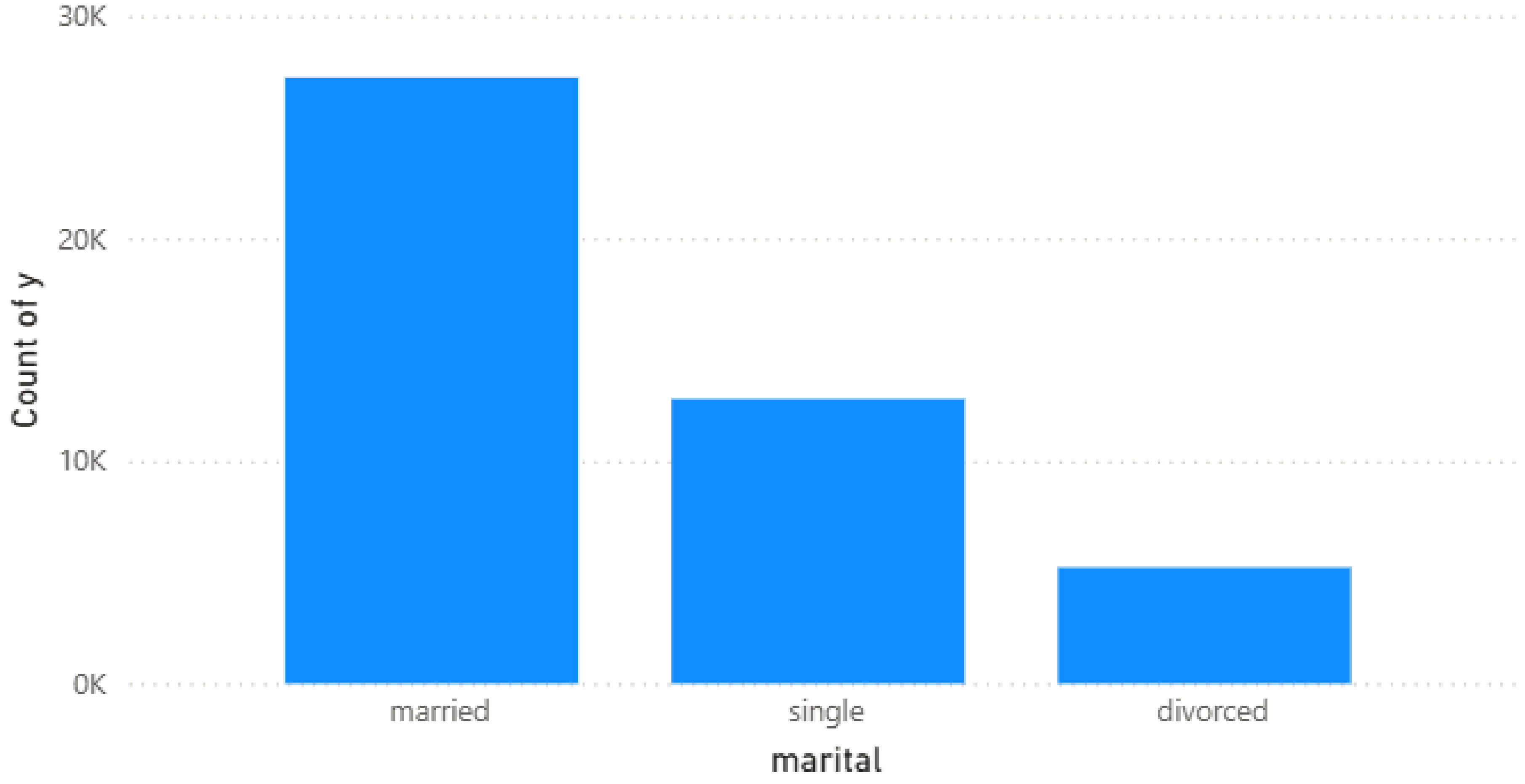
Count of job by job



Count of y by education



Count of y by marital



Key Deliverables of the Project

Summary of Outputs and Resources Created

Comprehensive Code Notebooks

The project includes well-documented **code notebooks** that detail the entire workflow from data preparation to model deployment. These notebooks serve as a reference for replicating analyses and can be easily shared with team members for collaborative efforts.

Trained Predictive Models

We have developed multiple **trained models** tailored for subscription prediction using various machine learning algorithms. These models are rigorously tested and validated to ensure they deliver reliable predictions for future marketing campaigns, enhancing targeting efficiency.

Interactive Dashboard and Reports

An **interactive dashboard** was created to visualize key performance metrics and insights derived from the project. This dashboard allows stakeholders to monitor campaign effectiveness and make data-driven decisions, encapsulating all relevant findings and recommendations in an accessible format.

Risks and Mitigation Strategies in the Project

Identifying and Addressing Key Challenges

Class Imbalance Risks

Class imbalance in subscription prediction can lead to biased models, favoring prevalent classes. To mitigate this, techniques such as oversampling minority classes, undersampling majority classes, or using synthetic data generation can be employed to enhance model performance and accuracy.

Data Leakage Concerns

Data leakage occurs when information from outside the training dataset is used inappropriately during model training, resulting in overly optimistic performance estimates. Implementing strict validation protocols and ensuring that data splits are performed correctly can help prevent this issue and maintain model integrity.

Privacy and Interpretability Challenges

Maintaining customer privacy is crucial, especially in banking campaigns. Employing techniques such as differential privacy and ensuring compliance with regulations like GDPR will safeguard customer information while preserving model interpretability to foster trust and clarity in decision-making processes.

Final Thoughts on Our Project

The conclusion of our project presents a **deployment-ready**, interpretable model pipeline that emphasizes strong ROI.
Key highlights include:

- Successful integration of machine learning and causal inference methodologies.
- Enhanced targeting of persuadable customers through uplift modeling.
- Development of a user-friendly dashboard for business insights.

Within customer relationship management (CRM) systems to streamline processes and enhance user experience.

References

1. S. Moro, P. Cortez, and P. Rita, “A data-driven approach to predict the success of bank telemarketing,” Decision Support Systems, vol. 62, pp. 22–31, 2014.

2. P. Burez and D. Van den Poel, “Handling class imbalance in customer churn prediction,” Expert Systems with Applications, vol. 36, no. 3, pp. 4626–4636, 2009.

3. European Central Bank, “Statistical Data Warehouse: Economic Indicators.” [Online]. Available: https://data.ecb.europa.eu

4. J. Neslin, S. Gupta, W. Kamakura, and D. Lu, “Defection detection: Measuring and understanding the predictive accuracy of customer churn models,” Journal of Marketing Research, vol. 43, no. 2, pp. 204–211, 2006.