

WELCOME TO THE FUTURE OF WORK WITH ENTERPRISE AI

Analyzing Customer Attrition with ChatGPT

An Enterprise AI Case Study

ABSTRACT

In this case study, we illustrate a scenario where a business analyst in a marketing department employs ChatGPT, acting as a virtual data scientist, to construct a model for analyzing and predicting customer attrition in a hypothetical company.

The entire conversational process, which includes generating synthetic data, running Python code in the background, and performing an in-depth analysis of statistical models, was completed in less than five minutes.

This case study can be easily understood without the prerequisite of programming, AI/ML, or data science skills.



TABLE OF CONTENTS

2	INTRODUCTION
3	PAINTING THE PICTURE: SCENARIO DESCRIPTION
4	SECTION ONE: CREATING SYNTHETIC DATASET
6	SECTION TWO: SEEK ADVICE ON THE ANALYSIS OF SYNTHETIC DATA
8	SECTION THREE: DATA EXPLORATION
10	SECTION FOUR: FEATURE ANALYSIS
12	SECTION FIVE: MODELING
14	RECAP
15	ABOUT THE AUTHOR



INTRODUCTION

Companies are beginning to realize that ChatGPT offers robust data science capabilities that can be harnessed by business analysts across a myriad of predictive and analytical evaluations. These evaluations span across diverse business use cases, including customer churn, as well as various forecasting scenarios such as enterprise cash flow or staffing within professional services companies.



It is important to emphasize that utilizing ChatGPT for comprehensive business data analysis and trend prediction doesn't necessitate expertise in complex fields like data science, machine learning, statistics, or Python programming.

This is because ChatGPT automatically integrates the principles and methods of these disciplines as needed.

In order to effectively leverage the power of ChatGPT, what business analysts need is a thorough grasp of the business features relevant to their objectives (such as predictive forecasting), which will enable more effective analyses and predictions. For instance, a market

business analyst should be well-versed in concepts related to customer churn (e.g., churn determinants, customer segmentation, and customer lifetime value (CLV) analysis).

In this case study, we demonstrate how a business analyst from a marketing department uses Enterprise ChatGPT, acting as a virtual data scientist, to build a model. This model aims to analyze and predict customer churn within a hypothetical telecom company.

- The methodology detailed in this case study can be utilized for all relevant business scenarios across various industries.
- Remarkably, the preparatory work detailed in this case study - including defining customer attrition, identifying relevant data points, reports, methods, and ChatGPT Prompts - was completed in under three days. Typically, such a task would take between eight to ten weeks and would involve specialists like a market analyst, a data scientist, and an IT solution designer or system analyst.
- The entire conversational process, which includes generating synthetic data, running Python code in the background, and performing an in-depth analysis of statistical models, was completed in less than 5 minutes.
- The model and accompanying information produced by ChatGPT can be compiled into a structured document. This document can then serve as a “business requirements” guide for your organization's AI or IT team, assisting them in implementing it within the enterprise AI environment and operationalizing it for related processes.

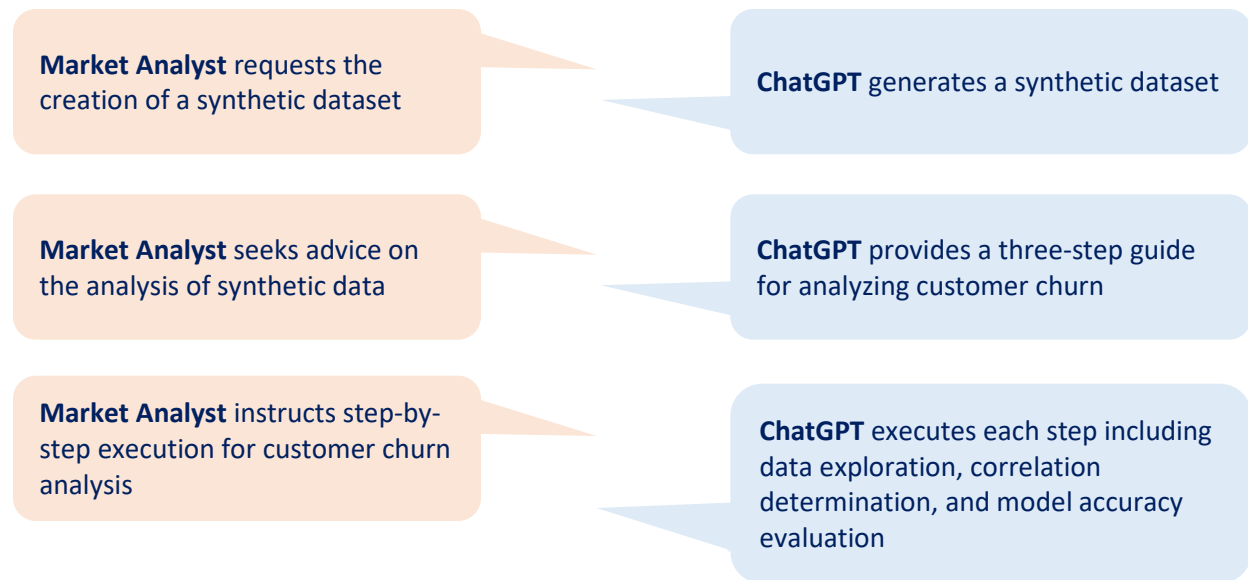
A synthetic dataset generated by ChatGPT is utilized for this experiment. It's crucial to note that real-world datasets containing actual customer data should never be used with the generic version of ChatGPT, in compliance with standard privacy and security regulations of your organization. By adhering to your company's AI governance policies, you can replicate synthetic models created using the generic version of ChatGPT in your secured enterprise AI platform.



PAINTING THE PICTURE: SCENARIO DESCRIPTION

As illustrated in the following chart, ChatGPT is acting as a data scientist, responding to the prompts provided by the market analyst. The conversation initiates with a market analyst from a telecom company prompting ChatGPT to generate a synthetic dataset relevant to customer churn analysis. Once the synthetic dataset is ready, the analyst prompts ChatGPT to outline the requisite steps for processing the dataset for churn analysis and forecasting future trends. Following the recommended steps provided by ChatGPT, the analyst instructs it to execute each step. In the final step, ChatGPT delivers key insights that include the results of the analysis, predictions of future trends, and suggestions for enhancing the models using more sophisticated algorithms and further refinements.

Outline of Dialogue Between Market Analyst and ChatGPT





SECTION ONE CREATING SYNTHETIC DATASET

In this section, the market analyst prompts ChatGPT to generate a synthetic dataset for customer churn analysis. In response, ChatGPT utilizes Python to establish and fill the most pertinent data fields associated with customer churn analysis with synthetic data. It then showcases a sample of the synthetic data alongside the Python code utilized in its creation.



Market Analyst

You are a data scientist. Generate a simple synthetic dataset to analyze and predict customer churn in telecommunications industry.



ChatGPT

Creating a synthetic dataset tailored for predicting customer churn in the telecommunications industry involves simulating data that mirrors real-world patterns, behaviors, and characteristics of customers who might decide to leave a service provider. While I don't have a pre-made dataset available, I can guide you on how to construct a basic synthetic dataset using Python. This dataset could include features commonly associated with churn, such as contract length, monthly charges, service usage, customer satisfaction levels, and demographic information.

Would you like a demonstration on how to generate a simple synthetic dataset for this purpose using Python?



Market Analyst

Yes.



ChatGPT

I've created a simple synthetic dataset with the following features that could be relevant for predicting customer churn in the telecommunications industry:

- Age: The customer's age.
- ContractLength: The length of the customer's contract in months (12, 24, or 36 months).
- MonthlyCharges: The amount charged to the customer each month.
- TotalCharges: The total amount charged to the customer over the lifetime of their account.
- DataUsage: The average data usage per month in GB.



- CustomerServiceCalls: The number of customer service calls made by the customer.
- PlanType: The type of plan the customer has subscribed to (Basic, Plus, or Premium).
- InternationalPlan: Whether the customer has an international plan (1 for yes, 0 for no).
- Churn: Whether the customer has churned (1 for yes, 0 for no).

Here's a preview of the first few rows of the dataset:

Customer ID	Age	ContractLength	MonthlyCharges	TotalCharges	DataUsage	CustomerServiceCalls	PlanType	InternationalPlan	Churn
1	56	24	69.92	7731.78	9.29	4	Premium	0	0
2	69	12	45.23	6112.2	1.36	5	Basic	1	1
3	46	12	129.24	7611.1	3.62	7	Plus	0	0
4	32	24	122.69	5650.07	2.87	6	Basic	0	0
5	60	12	83.99	2454.61	3.14	7	Premium	0	0

This synthetic dataset can be used for exploratory data analysis, feature engineering, and training machine learning models to predict customer churn. Adjustments to the distributions, such as changing the probabilities for plan types or churn, can help simulate different scenarios and customer behaviors.

The snippet of the Python code that ChatGPT created:

```
#Python code created by ChatGPT
#Generating a synthetic dataset based on 1000 samples.
import pandas as pd
import numpy as np

# Seed for reproducibility
np.random.seed(42)

# Number of samples
n_samples = 1000

# Generate synthetic data
data = {
    'CustomerID': range(1, n_samples + 1),
    'Age': np.random.randint(18, 70, n_samples),
```




SECTION TWO SEEK ADVICE ON THE ANALYSIS OF SYNTHETIC DATA

In this section, the market analyst prompts ChatGPT for guidance on how to investigate the underlying reasons for customer churn. Responding to this prompt, ChatGPT outlines a comprehensive three-step process designed to dissect and understand the potential factors contributing to customer churn.



Market Analyst

Use the dataset that you created with Python to explain the root cause of customer churn.



ChatGPT

To analyze the root cause of customer churn using the synthetic dataset we created, we would typically follow a data science workflow that involves data exploration, feature analysis, and modeling. However, since our dataset is synthetic and randomly generated, any insights or conclusions we draw might not accurately reflect real world phenomena. Nevertheless, I can guide you through a hypothetical analysis to illustrate how such a process might be conducted on an actual dataset:

- Step 1: Data Exploration
- Step 2: Feature Analysis
- Step 3: Modeling

After identifying key features, we'd use machine learning models to predict churn and identify the most important features contributing to customer churn. Common models for this type of analysis include logistic regression, decision trees, and random forests. These models can provide insights into which features are most predictive of churn, helping to identify the root causes.

Hypothetical Analysis Based on Synthetic Data

Given our synthetic dataset, let's hypothesize that after performing exploratory data analysis and modeling, we found the following factors to be significantly associated with customer churn:

- High Monthly Charges
- Short Contract Length
- High Number of Customer Service Calls
- Lack of an International Plan



Conclusion

In a real-world scenario, these insights could inform strategies to reduce churn, such as adjusting pricing, improving customer service, offering more flexible international plans, or providing incentives for longer contract commitments.

Remember, the above analysis is purely hypothetical, given the synthetic nature of our dataset. Real-world data would require thorough cleaning, exploration, and validation to derive accurate insights.

The snippet from ChatGPT's comprehensive explanation of the approach used to analyze synthetic data:

Step 1: Data Exploration

First, we'd explore the dataset to understand the distribution of features, the balance between churned and retained customers, and any immediate correlations visible between features and the churn rate.

Step 2: Feature Analysis

Next, we'd analyze how different features correlate with customer churn.

Step 3: Modeling

After identifying key features, we'd use machine learning models to predict churn and identify the most important features contributing to customer churn.

The snippet of the Python code that ChatGPT created:

```
#Python code created by ChatGPT
# Step 2: Feature Analysis - Correlation with Churn
# Calculate correlations between all numerical features and churn
correlation_matrix = df.corr()

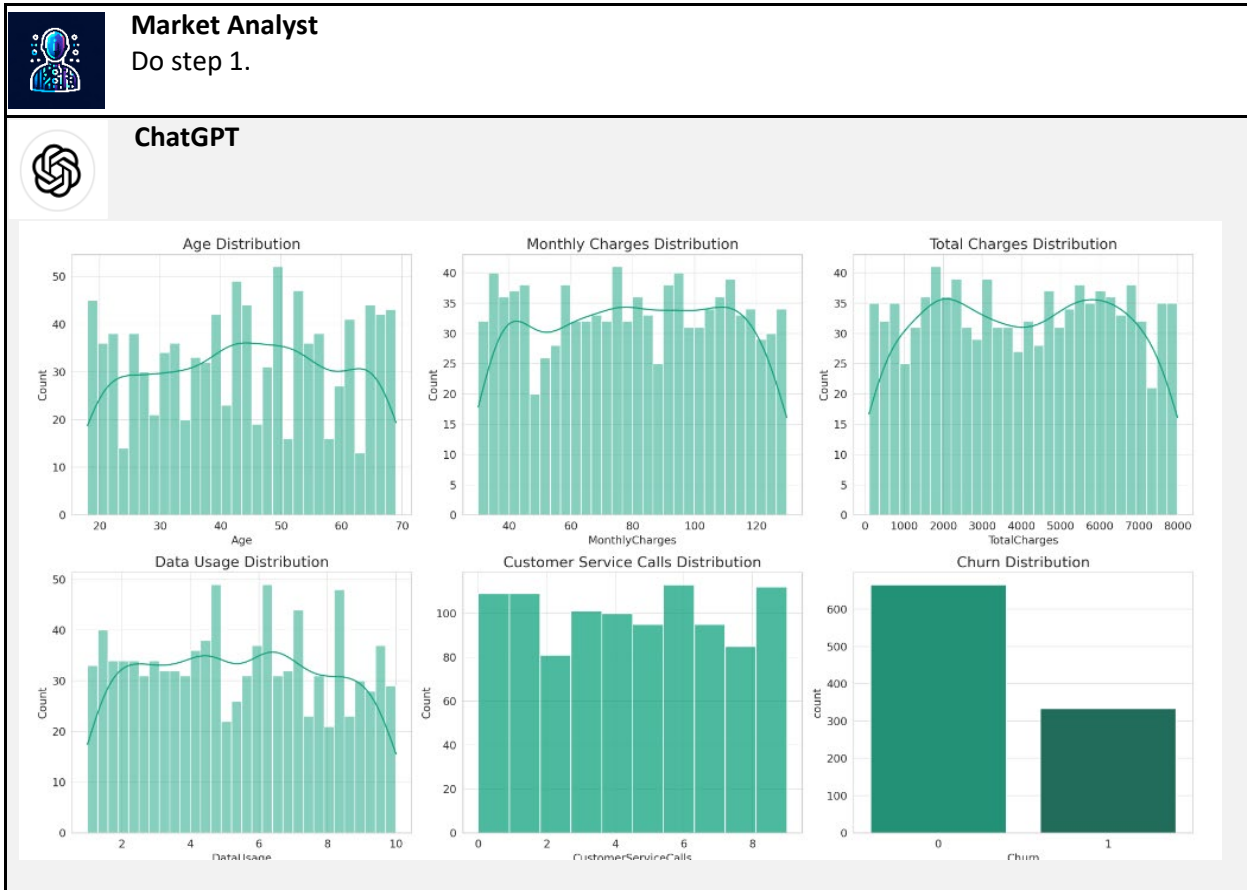
# Visualize the correlation matrix with a heatmap
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f")
plt.title('Correlation Matrix')
plt.show()
```




Section Three

DATA EXPLORATION

In this section, the market analyst instructs ChatGPT to carry out the first step of the churn analysis process previously suggested. ChatGPT undertakes a "data exploration" task, offering insights underpinned by statistical summaries and visual representations of the synthetic data fields. The Python code provides the libraries that were utilized for statistical functions and visual representations.



From our initial data exploration and visualization, we can observe the following about the synthetic dataset:



Summary Statistics

- **Age:** The customers' ages range from 18 to 69 years, with a mean age of approximately 43.8 years.
- **Contract Length:** Contract lengths are distributed among 12, 24, and 36 months, with a mean around 23.7 months, indicating a mix of short-term and long-term customers.
- **Monthly Charges:** Monthly charges vary widely from \$30.00 to \$129.82, with an average monthly charge of around \$79.94.
- **Total Charges:** Total charges to customers range from \$100.09 to \$7991.25, with a mean total charge of approximately \$4038.37, suggesting a wide variance in customer lifetime value.
- **Data Usage:** Average monthly data usage is distributed between 1 GB and nearly 10 GB, with a mean of about 5.41 GB.
- **Customer Service Calls:** The number of customer service calls ranges from 0 to 9, with an average of 4.48 calls per customer.

Visualizations

- **Age Distribution:** The age distribution is fairly uniform, indicating a diverse range of customer ages.
- **Monthly Charges Distribution:** Monthly charges are uniformly distributed across the range, showing that customers are on a variety of plan pricing.
- **Total Charges Distribution:** Total charges are spread across a wide range, indicating variability in how long customers have been with the provider and how much they have been charged over time.
- **Data Usage Distribution:** Data usage is evenly distributed, suggesting that customers have varied data needs.
- **Customer Service Calls Distribution:** The distribution of customer service calls shows that while many customers have made few calls, there's a wide spread in the number of calls made.
- **Churn Distribution:** Approximately 33.4% of the customers have churned, and 66.6% have stayed. This indicates a significant churn rate that would require further analysis to understand the underlying causes.

Initial Observations

The initial exploration provides a broad overview of the dataset's characteristics. To understand the root causes of customer churn, we would next look into correlations between these features and churn. For instance, high monthly charges, a high number of customer service calls, or short contract lengths could potentially be linked to higher churn rates. Detailed analysis using statistical tests and predictive modeling would help identify the most significant predictors of churn in this dataset.

Remember, the insights derived here are based on synthetic data and are for illustrative purposes. Analysis of a real dataset might reveal different patterns and require different approaches to uncover the true drivers of customer churn.

The snippet of the Python code that ChatGPT created:

```
#Python code created by ChatGPT
#Data Exploration and Visualization
# Re-import necessary libraries and re-generate the dataset due to reset
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```



SECTION FOUR FEATURE ANALYSIS

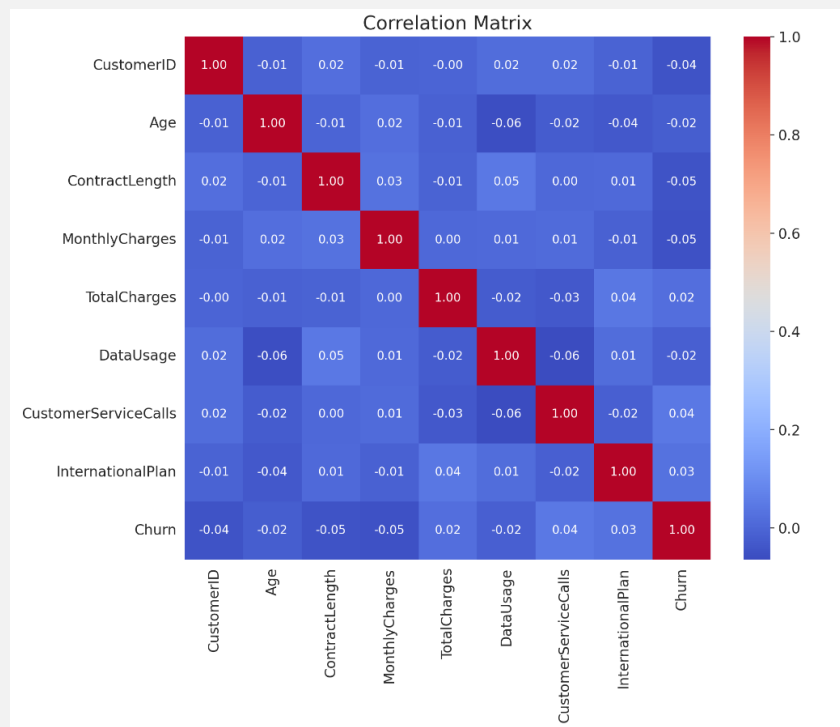
In this section, the market analyst directs ChatGPT to execute the second step of the previously recommended churn analysis process. ChatGPT runs a "feature analysis" task, presenting a correlation matrix that depicts the relationships between the synthetic data fields. The accompanying Python code specifies the libraries used for statistical calculations and visual displays.



Market Analyst
Do step 2.



ChatGPT





The correlation analysis between various features and customer churn reveals the following insights:

- **Churn** has a correlation coefficient of 1 with itself, as expected.
- **Customer Service Calls** show a slight positive correlation (0.045) with churn, suggesting that customers who make more service calls are slightly more likely to churn. This could indicate dissatisfaction with the service.
- **International Plan** has a small positive correlation (0.028) with churn, implying that having an international plan might slightly increase the likelihood of churn, possibly due to expectations not being met.
- **Total Charges** also show a slight positive correlation (0.022) with churn. This is interesting because one might expect customers who have spent more with the company over time to be less likely to churn. However, this correlation is very weak.
- **Data Usage** and **Age** show very small negative correlations with churn (-0.018 and -0.019, respectively), suggesting that higher data usage and older age could marginally reduce the likelihood of churn, though the effect is minimal.
- **CustomerID** has a negative correlation (-0.045) with churn, which is likely coincidental and not meaningful, as CustomerID is just an identifier.
- **Contract Length** and **Monthly Charges** have negative correlations (-0.047 and -0.050, respectively) with churn, indicating that longer contracts and lower monthly charges might slightly decrease the likelihood of churn. This suggests that customers on longer contracts and/or paying less per month are less inclined to leave, possibly due to perceived value or the inconvenience of switching providers.

Key Takeaways

While the correlations are generally weak, indicating no single strong predictor within this set of features, the analysis points towards a few potential areas of interest:

- **Customer Service Calls:** A higher number of calls might be an indicator of customer dissatisfaction. Improving customer service and resolving issues efficiently could potentially reduce churn.
- **Contract Length:** Longer contracts appear to have a modest protective effect against churn. This could be leveraged by encouraging customers to commit to longer terms through incentives or better pricing.
- **Monthly Charges:** Lower monthly charges are slightly associated with reduced churn. This suggests pricing strategy and perceived value are important factors in customer retention.

It's important to note that correlation does not imply causation, and these findings would need to be further investigated, ideally through a combination of more detailed statistical analysis and predictive modeling. Additionally, given that this dataset is synthetic, real-world data could reveal different patterns and relationships.

The snippet of the Python code that ChatGPT created:

```
#Python code created by ChatGPT
# Step 2: Feature Analysis - Correlation with Churn
# Calculate correlations between all numerical features and churn
correlation_matrix = df.corr()
# Visualize the correlation matrix with a heatmap
plt.figure(figsize=(10, 8))
sns.heatmap(correlation_matrix, annot=True, cmap='coolwarm', fmt=".2f")
```




SECTION FIVE MODELING

In the final section, the market analyst directs ChatGPT to carry out the third stage of the churn analysis process that was previously suggested. ChatGPT undertakes a "modelling" task, selecting the random forest algorithm and clarifying the significance of the statistical output in terms of accuracy. Additionally, it delivers key insights about the overall model accuracy and outlines potential improvements for deploying the model with real-world data. The provided Python code details the libraries employed for statistical computations.



Market Analyst
Do step 3.



ChatGPT

The Random Forest model achieved an accuracy of 64% on the test set, which is a basic level of performance for this synthetic dataset. Here's a breakdown of the model's evaluation metrics:

- **Precision** for class 0 (non-churned customers) is 0.69, meaning that when the model predicts a customer will not churn, it is correct about 69% of the time.
- **Recall** for class 0 is 0.86, indicating that the model successfully identifies 86% of the actual non-churned customers.
- **F1-score** for class 0 is 0.77, which is a balance between precision and recall, suggesting a relatively good performance for this class.

For class 1 (churned customers), the performance metrics are significantly lower:

- **Precision** is 0.32, **Recall** is 0.15, and **F1-score** is 0.20. This indicates that the model struggles to accurately identify churned customers, often misclassifying them.

Key Takeaways

- The model's overall accuracy of 64% shows there is room for improvement. Given the relatively high recall for non-churned customers but low recall for churned customers, the model is more conservative and better at identifying customers who will stay rather than those who will churn.
- The imbalance in performance metrics between the two classes suggests that the model could benefit from techniques aimed at addressing class imbalance, such as oversampling the minority class or adjusting class weights in the model.
- Further feature engineering, including encoding categorical variables like 'PlanType' and possibly creating new features, could also improve the model's performance.



- Since this is a synthetic dataset, real-world datasets with actual customer behavior patterns may yield different and potentially more insightful results.

Improving the model could involve experimenting with different algorithms, tuning hyperparameters, and employing more sophisticated techniques to handle imbalanced datasets.

The snippet of the Python code that ChatGPT created:

```
#Python code created by ChatGPT
# Step 3: Determining model pertinence and precision

from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score, classification_report

# Preparing the data for modeling
# Dropping 'CustomerID' as it's not a useful feature for prediction
# PlanType is categorical and needs encoding before use
X = df.drop(['CustomerID', 'Churn', 'PlanType'], axis=1)
y = df['Churn']
```




RECAP

Our case study (exemplified by ChatGPT) presents an exploration of predictive and generative AI in the enterprise to help the reader visualize how the domain of data science can be democratized within the sphere of business analysis. By leveraging ChatGPT or similar copilots, acting as a virtual data scientist, we've demonstrated a practical and efficient approach to tackling customer attrition analysis—a critical challenge in marketing.

This innovative methodology not only simplifies the process of data analysis but also significantly reduces the time and expertise barrier traditionally required for such tasks.

Our ability to prepare the scenario in just three days and run it in less than five minutes, without requiring direct programming, AI/ML expertise, or data science skills, signifies a transformative change in how businesses can set requirements for data-driven decision-making.

This case study demonstrates the potential to empower knowledge workers across departments, enabling them to perform complex analyses with ease and make informed decisions based on the insights derived.

From an economic impact perspective, McKinsey and other economists estimate that AI could generate trillions of dollars in revenue. This comes from new business opportunities, improved efficiency, automation of routine knowledge-driven tasks, and the creation of entirely new industries centered around AI. **It's a gold rush scenario, with businesses racing to harness the power of AI and reap its benefits.**

Many companies understand the potential value of generative AI but are struggling with getting started. Blue Monday is already helping organizations develop customized roadmaps that outline and deliver on short-, medium- and long-term opportunities for capitalizing on this change in the business landscape:

- **Identifying Impactful Use Cases:** Collaborating with you to pinpoint areas where Enterprise AI can **supercharge your business**. Analyzing workflows, data landscape, and strategic goals to identify high-impact applications in the areas of data-driven predictions.
- **Seamless Integration and Optimization:** Ensuring seamless integration of generative Large Language Models (LLMs) into existing infrastructure, data pipelines, and applications.
- **Building AI Fluency Across Teams:** We don't believe in black-box solutions. We provide **tailored training and support** to empower your team to understand and utilize Enterprise AI tools effectively.
- **A Holistic Approach Beyond Enterprise AI:** As your trusted advisor, we offer a comprehensive suite of consulting services that extend beyond Enterprise AI. Our expertise across various enterprise systems, including Enterprise Resource Planning (ERP), Customer Relationship Management (CRM), and more, enables us to provide tailored advice and strategies that consider the specific nuances of each client's situation.

- It's important to note that AI adoption won't happen overnight or by adopting "do-it-yourself" approaches. Your business will need to strategically invest in Enterprise AI, upskill your workforce, and adapt your operating model to fully harness the benefits of AI.
- It's a complex journey, but with the right partner and guidance, your enterprise can successfully navigate this technological revolution and leverage the potential of Enterprise AI.
- Our partnership approach not only eases the transition but also ensures that your business is utilizing AI to its fullest potential, thereby getting the most out of your investment.



ABOUT THE AUTHOR

Siamak Razmazma is a seasoned strategic advisor specializing in technology transformation within enterprise initiatives. With a focus on Enterprise AI and conventional business systems, he is a respected thought leader and is known for articulating the business advantages of investing in new enterprise technologies. As a trusted advisor to leading companies across diverse industries, Siamak has helped numerous organizations stay competitive by developing future-proof business system strategies and facilitating transitions from legacy systems to latest technologies. With over three decades of experience under his belt, Siamak

presently leads Blue Monday Consulting as its founder and president.

Siamak fervently advocates for the potential of emerging technologies like predictive and generative AI to significantly enhance enterprise operations. Passionate about demystifying complex tech concepts, he aims to break them down into simpler ideas that can be understood and appreciated by those without deep technical background. Siamak holds an advanced doctoral degree in anthropological sociology from the University of Saint Denis, France, and a higher education diploma in computer studies from the University of Nanterre, France. This unique academic blend equips him with a distinctive viewpoint on enterprise business technologies.

Siamak can be reached at siamak@bluemonday.io.



GUEST CONTRIBUTOR

Philip Boken has spent his career helping companies design and leverage leading practices and technologies en route to the realization of processes that enable step change improvements to top and bottom-line performance. Philip has been a proponent of strategic investment of technology starting with the era of reengineering and the advent of email, the internet and ERP systems. While serving as an expatriate in Europe in the late 1990s, he led an international team that rolled

out the Business Performance Improvement methodology, training clients and consultants on the Balanced Scorecard and how to get ahead of the curve by investing in (and developing processes around) the then burgeoning technologies of eCommerce. Also of note is his experience helping clients leverage some of the first cloud-based SaaS tools for finance (such as BlackLine) and collaborating with the finance function of the first SaaS CRM company, helping them implement such tools within their organization.

Philip brings to the table strong depth and breadth across finance and operations, having worked with startups, middle market private equity portfolio companies, and public and private Fortune 100 enterprises. He has played a leading role in the development of finance and operations process frameworks, created techniques for diagnosing finance (notably the financial close) and has specific views on the importance of collaboration between finance and operations, especially in terms of performance measurement and analytics. In addition to serving as an interim executive for companies transitioning to new ownership, Philip contributes regularly as sounding board and mentor for companies seeking fresh ideas for how to approach challenging operational issues, including how and when to leverage enabling technologies. He'll often be the first to advise his clients that technology is not the immediate "silver bullet" solution, so that foundational issues can be addressed in order to enable a successful strategic investment in technology.

Philip can be reached at philip@pcbandco.net.