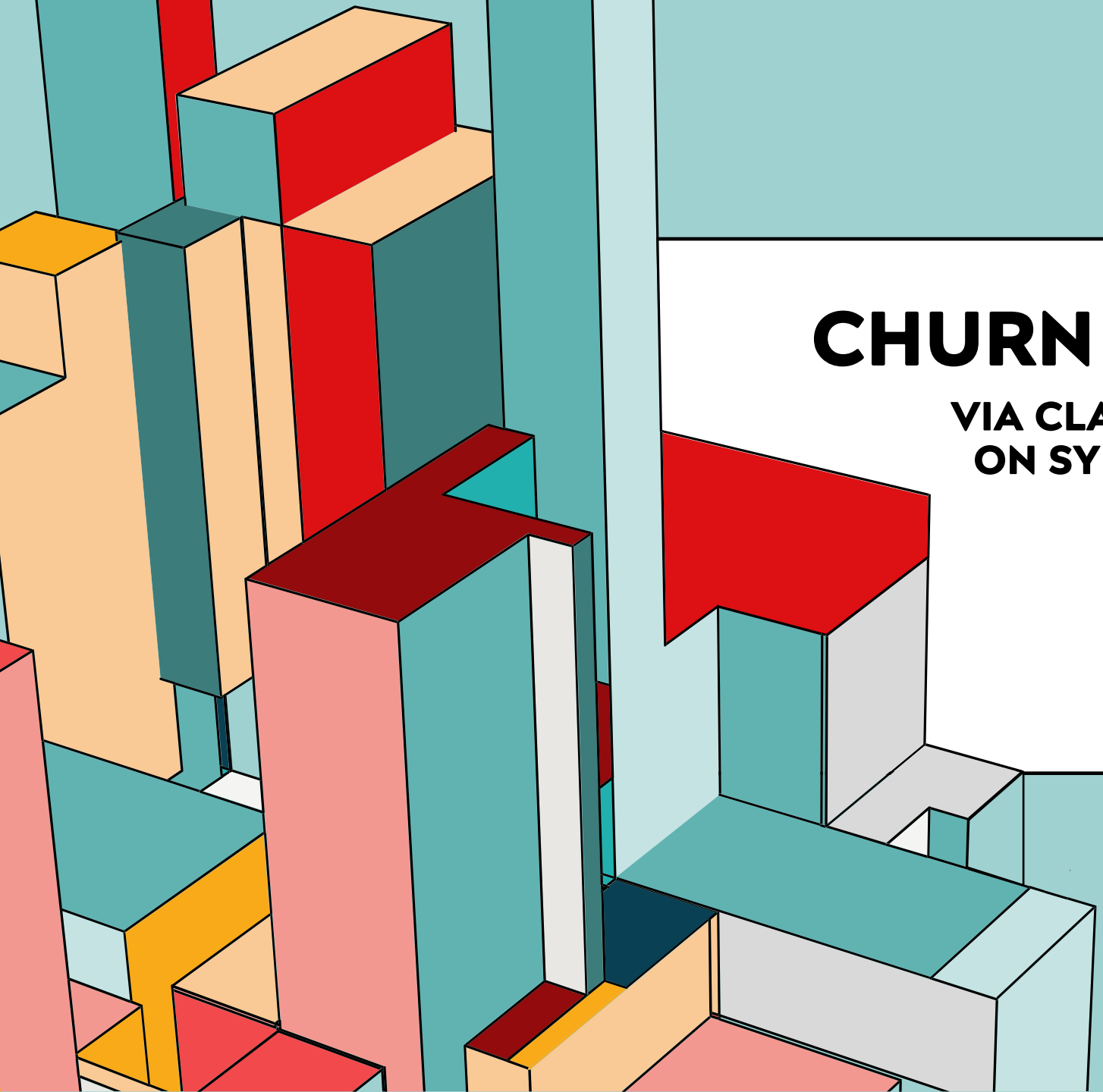


CHURN PREDICTION

VIA CLASSIFICATION ML MODELS ON SYRIATEL CHURN DATASET

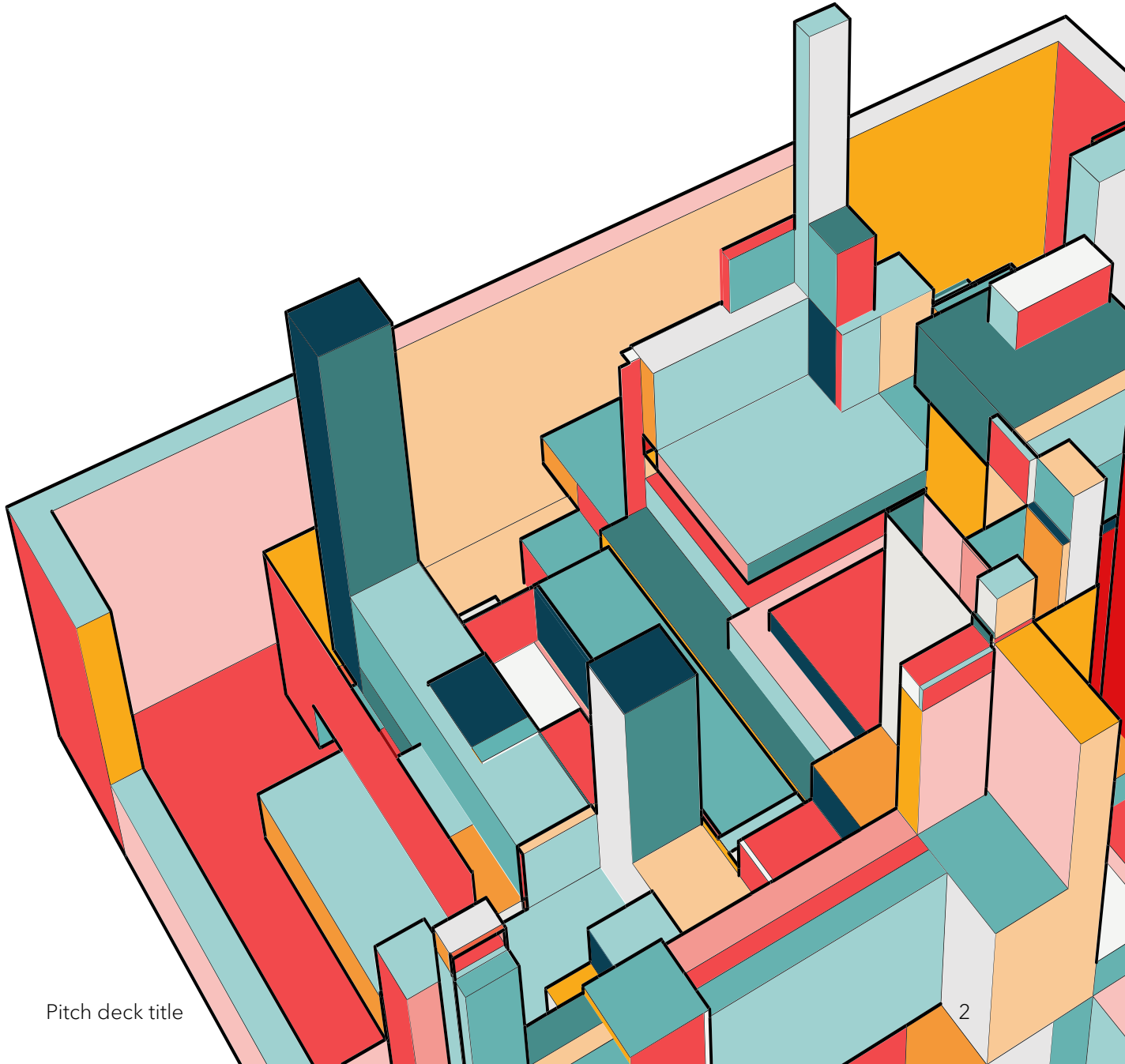
Chris Otieno

INTRODUCTION

I have analyzed the SyriaTel Customer Churn data to advise pricing decisions to be made by **BoraComm**, a telecommunications company.

Exploratory data analysis gave us insights into the data from which I derived the determinant of a customer's decision to stop using their services using various modeling techniques.

From this we are able to get three main business recommendations.



BUSINESS PROBLEM

PRICING CHALLENGES

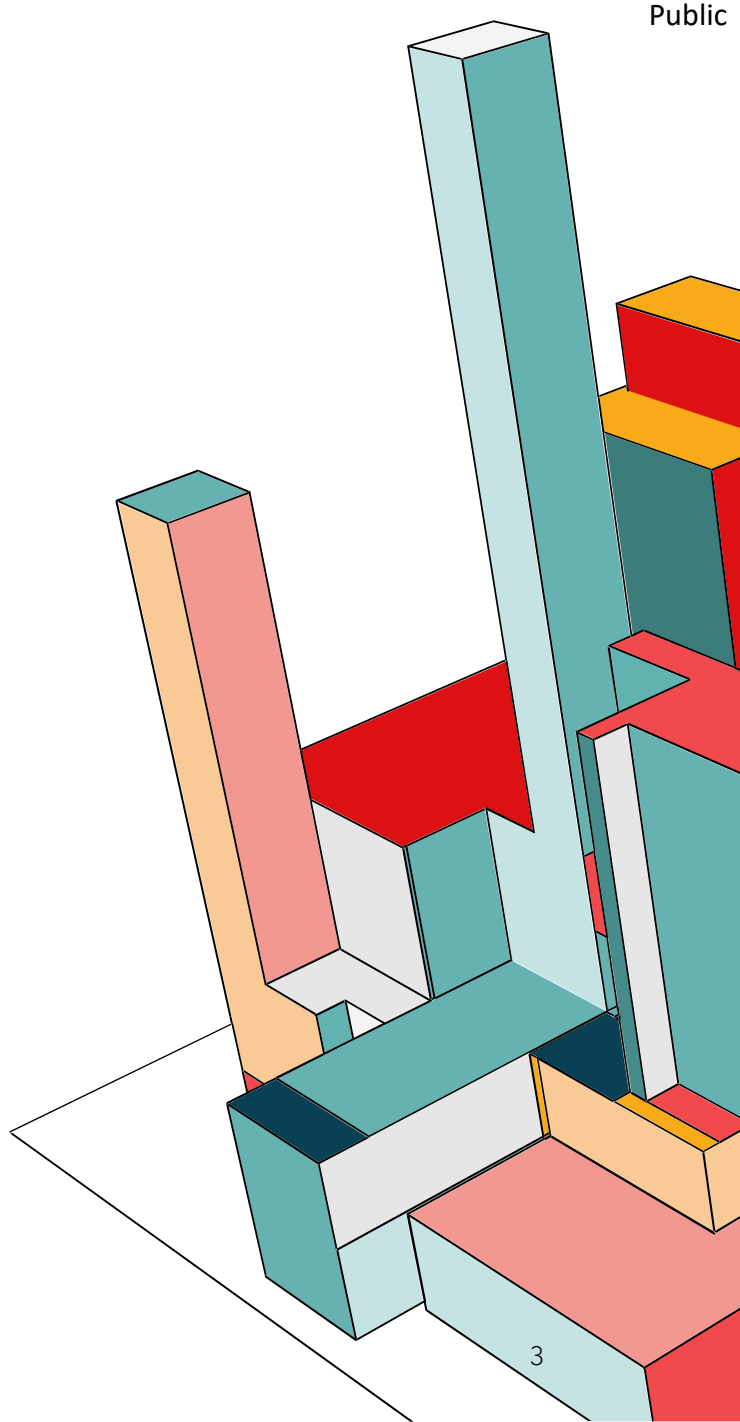
Even though telcos have a wealth of customer data available to them, they do not use it to predict which prices and products are most used by their customers

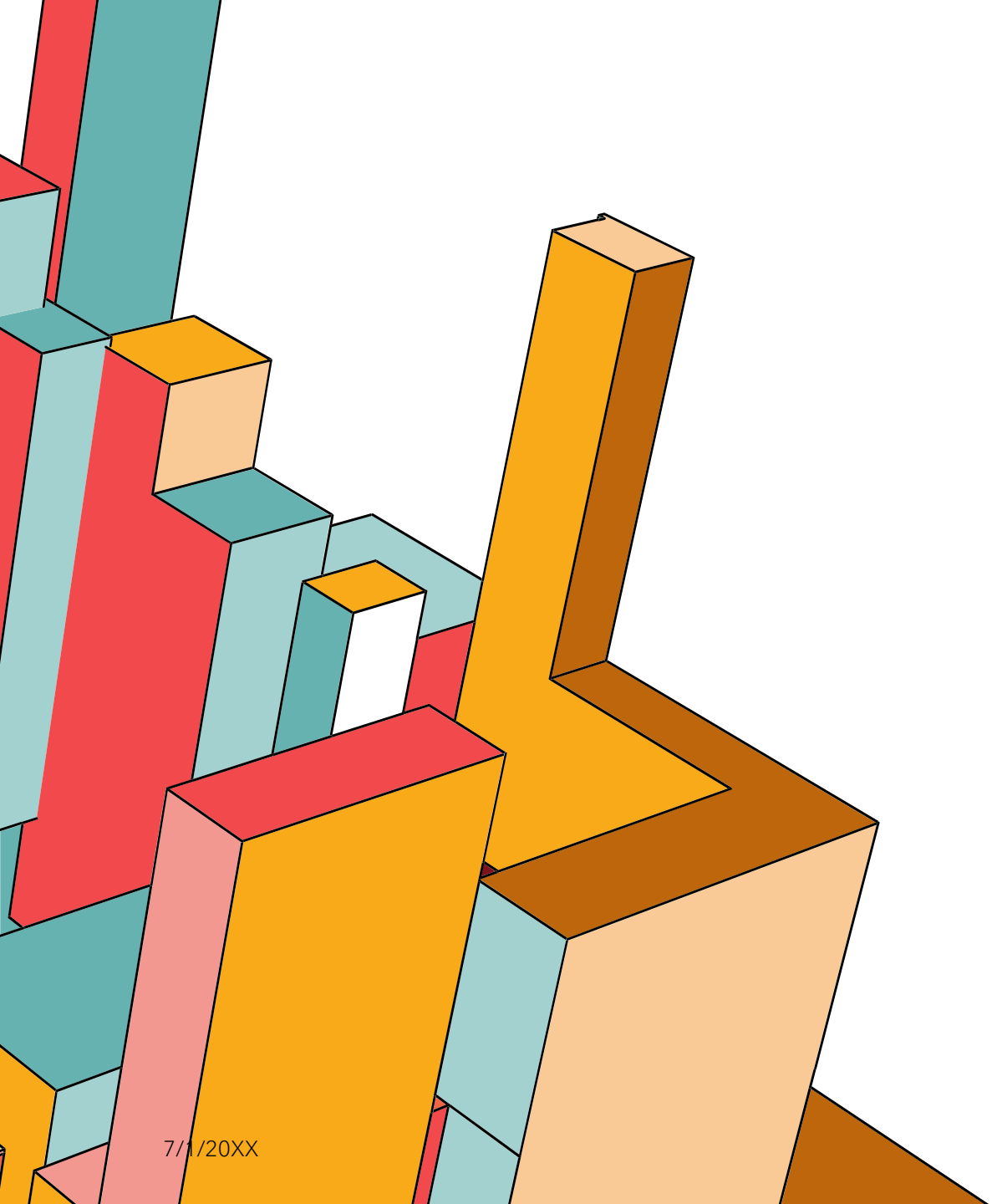
CUSTOMER RETENTION

There is a need to determine what factors cause a customer to abandon continuous use of a product in order attract the best buyers for the longest time

FINANCIALS

Supervised learning models accurately predict how loyal a customer will be to BoraComm's services





DATA UNDERSTANDING

SyriaTel Customer Churn dataset

The datasets contains valuable variables such as

- Account length
- International plan
- customer service calls
- Minutes spent on various types of calls
- Total charges on various types of calls

TARGET VARIABLE

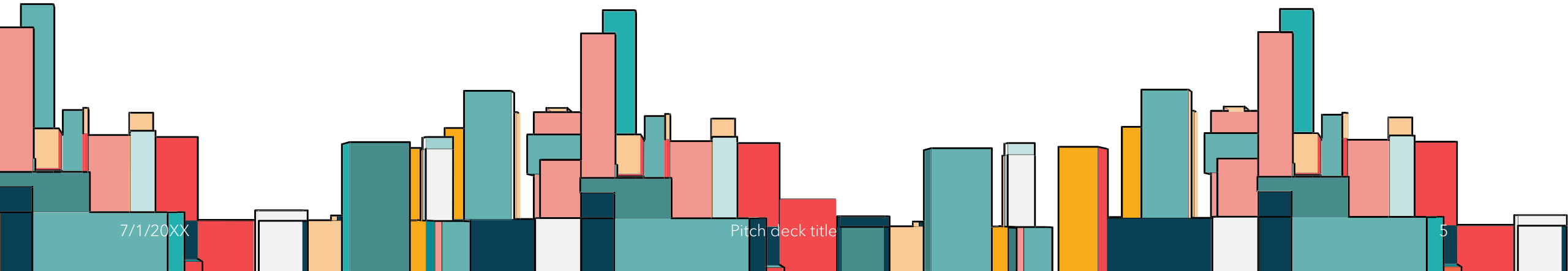
the **churn**, it serves as the independent variable in our prediction, we determine whether or not a customer has stopped using our services entirely

BUSINESS STAKEHOLDER

BoraComm

is a fictional company in the telecommunications.

They specialize in phone calls both locally and internationally with additional features such as voice mail.



DETERMINANTS

Total Minutes spent on **international calls** explains 23% likelihood to price

Total eve minutes

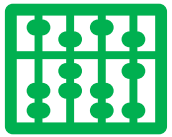
Number of **customer service call** increases likelihood to churn

Customer Service

Total charge of **day calls** leads to lower likelihood to churn

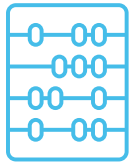
Total day charge

MODEL ACCURACY



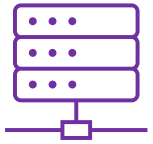
Naïve Bayes was the most inaccurate model at 0.8545

Naïve Bayes



Logistic Regression was the least inaccurate model at 0.8560

Logistic Regression



K-Nearest Neighbours was the second-best model at 0.8950

K-NN

MOST ACCURATE MODEL

Decision Trees

- **Decision Trees** was the most accurate model at 0.9235
  - Confusion Matrix was [[560 6] [64 37]]
    - **precision** is below 0.90

Insights:

- Likelihood to churn is tied directly to the number of customer service calls made
- Customers with frequent international calls least likely to churn

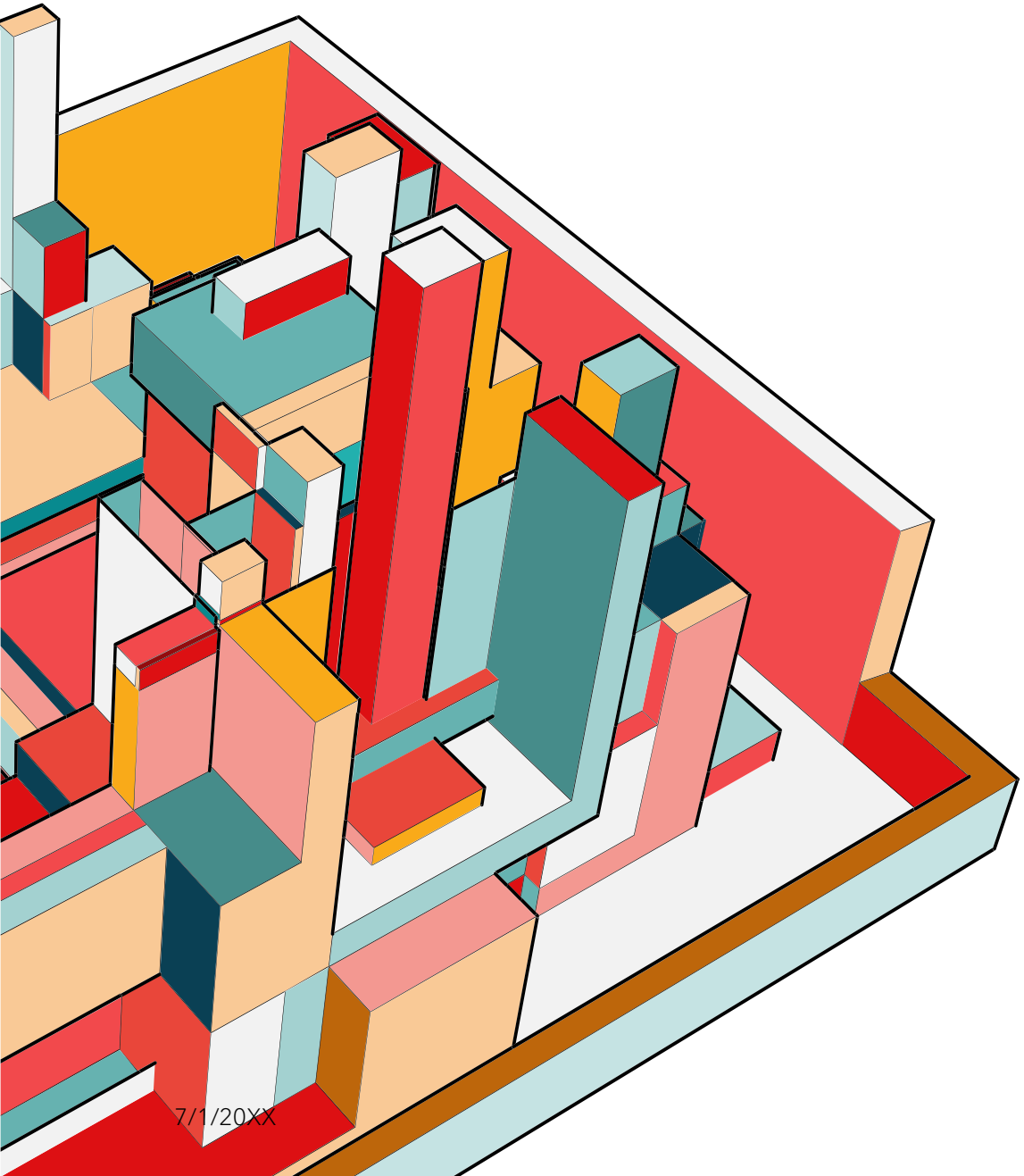
NEXT STEPS

EXPLORATORY DATA ANALYSIS

- More EDA on factors that might affect customer churn
- Our best performing model accounted for less than 70% of the variance in churn
- Further exploration should be done on factors not in the dataset that affect customer churn

DIRECT MARKET RESEARCH

- Direct market research from other players in the telecoms market
- Direct feedback will enhance the models by providing key dimensions missing in the datasets



CONCULSION

This prediction equips **BoraComm** with data-driven insights and predictive modeling strategies to make strategic decisions in the competitive telecommunications market.