

# Empowering Sales Management with Potential Detection And Conversion Analysis



MBA team: Hazel Sima, Peijun Xu | MIT Faculty: Dimitris Bertsimas  
Thermo Fisher Scientific: Benjamin Preston, Fabio Piccione, David Gieda, Tuan Ngo, Varun Shivkumar



## Project Overview



### Division Level: Transplant Diagnostic Division

#### Project I: Potential Detection

Objective: Identify potential sales quantity for two major products L & S by state

Goal: Quantify the sales opportunities to fine tune the sales structure and synergizes sales forces in regions with high potential

## Sales Team Management



### Division Level: Transplant Diagnostic Division

#### Project II: Conversion Analysis

Objective: Segment Customers into sales buckets based on conversion behaviors and provide customized marketing strategy recommendations

Goal: Generate product profile for each customer and put them into predictive buckets, suggesting future actions to research on how customers react to product updates



### Group Level: Special Diagnostic Group

#### Project III: Sales Forecast

Objective: Build Top-down Forecast model pipeline to predict future sales value on a monthly bases for 27 product groups

Goal: Provide accurate future revenue prediction and credibility to change sales actions with appropriate attention on accelerated markets

## Potential Detection

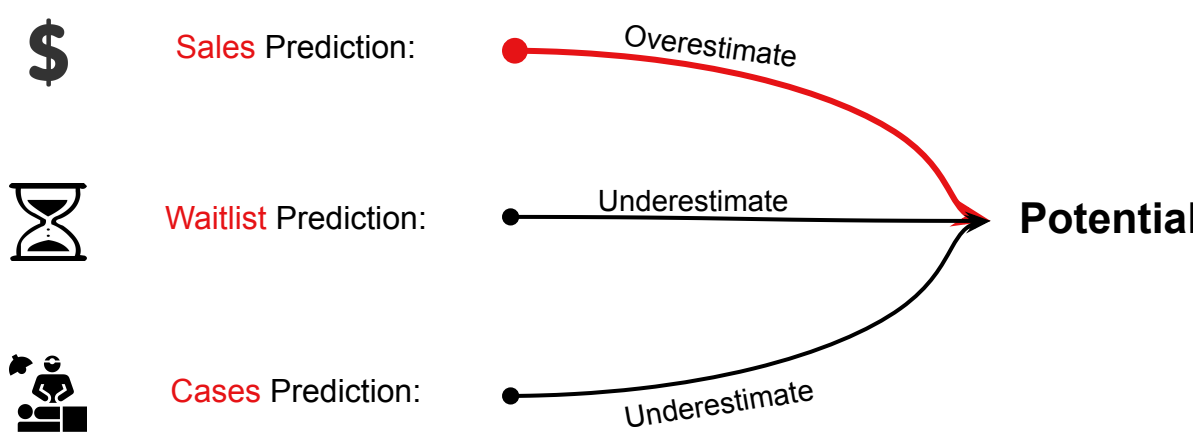
### Data

- UNOS Transplant Data from 1988 to 2021
  - Transplant cases performed yearly in each state
  - Transplant waitlist added yearly in each state
- Company Transactional Data
  - Quarterly revenue collected from each center from 2010 to 2021

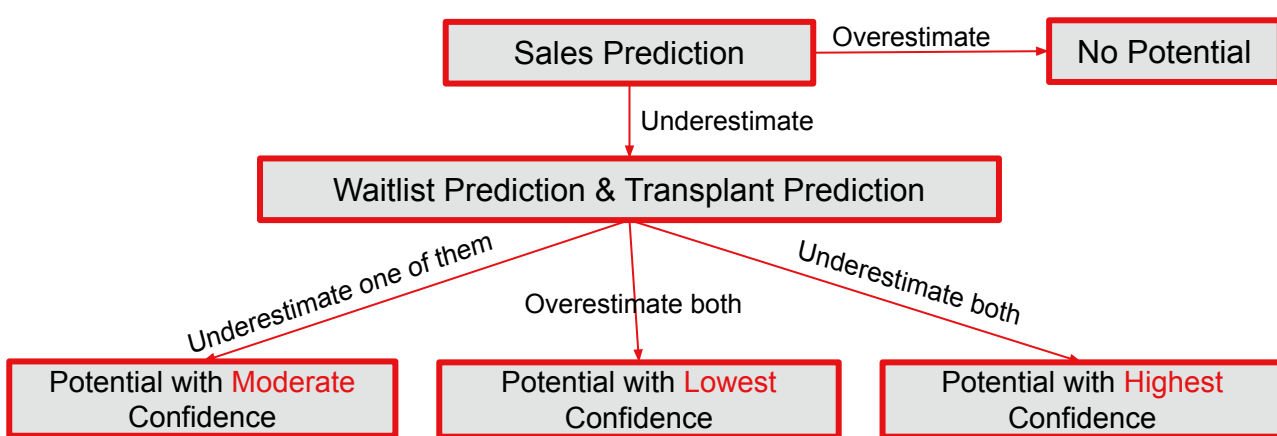


### Modeling

- We applied linear regression models to predict three quantities shown below. Sales prediction is the main indicator that quantifies potential and predictions for waitlists and cases help us better qualify our predictions

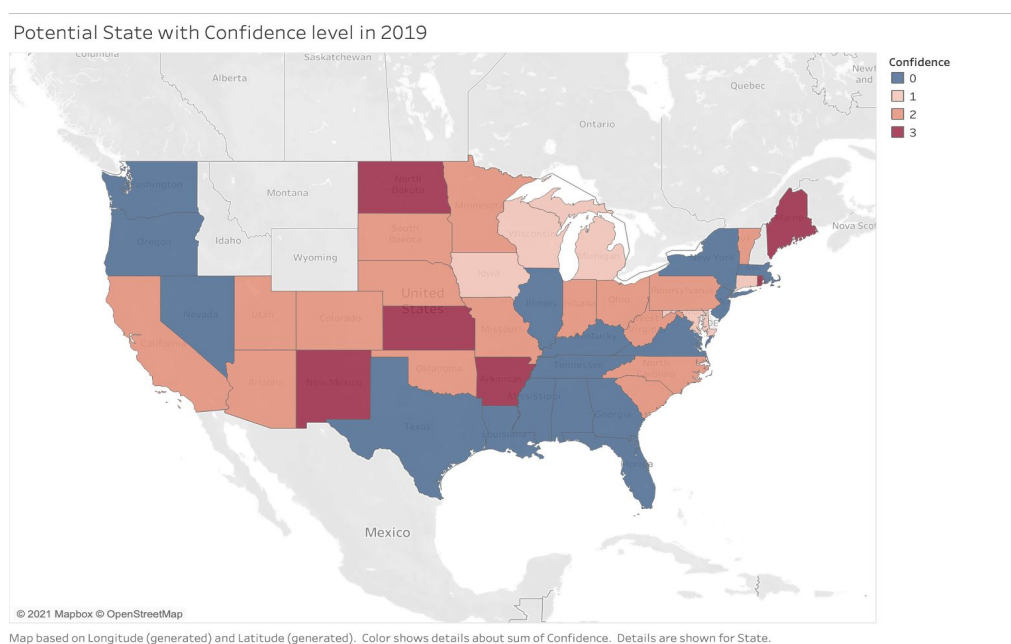


- Leverage three perspectives to generate confidence levels for prediction



### Results

- Identified Middlewest, Northeast and Southwest as the top 3 regions with a lot of potential and high confidence



## Conversion Analysis

### Data

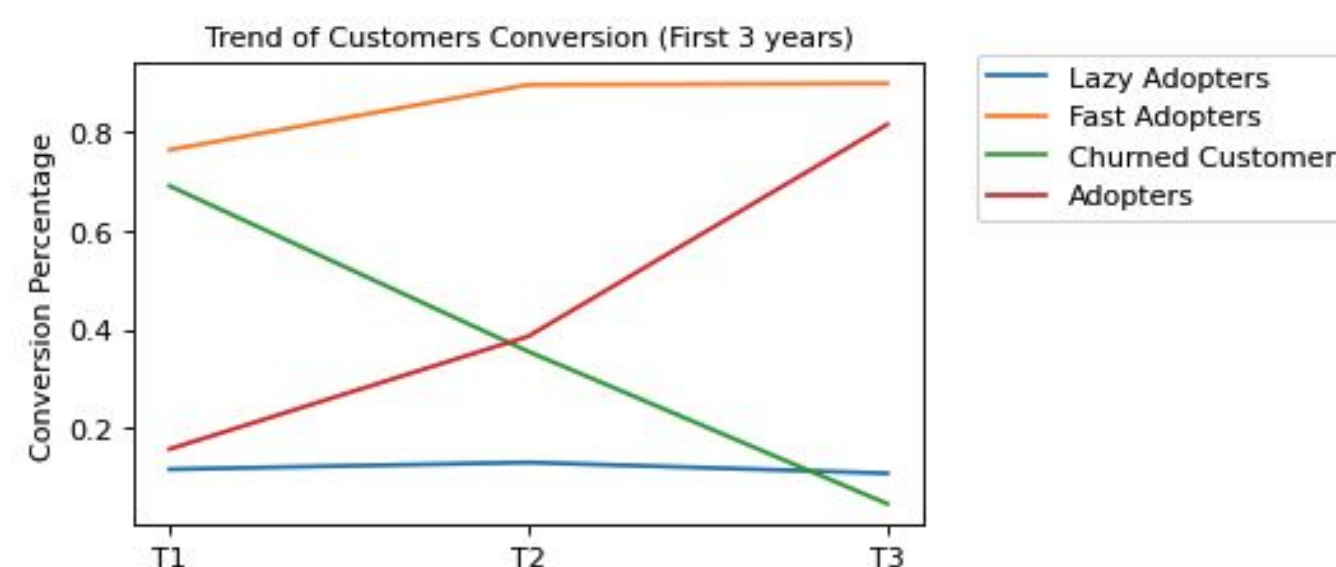
- Company Transactional Data
  - Quarterly sales quantity for an old version product and an updated version from 2018 to 2021 for individual customers

### Modeling

- Data Manipulation
  - We split the customer into three groups:
    - Customer who just purchased the old version product
    - Customer who just purchased the new version product
    - Customer who purchased both
  - We calculated the sum and difference of the two products sold by year
  - We calculated the conversion percentage time series data on the share of the new products over the total amount of products sold
- Method
  - Cluster on regression coefficient:** We fit linear regression models on the sum and difference sales quantity of the two products and cluster the customers based on the coefficients from the two models
  - Unsupervised Clustering-Hierarchical Clustering:** We clustered the customers in to different buckets based on the conversion percentage time series for three years

### Results

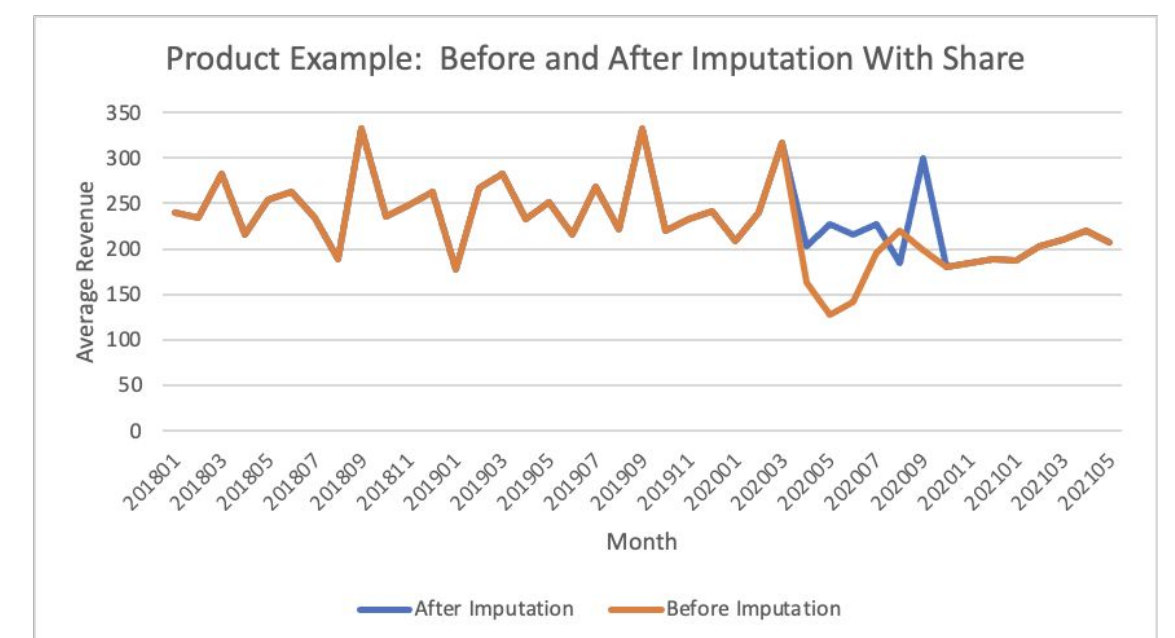
- We Identified four groups of customers with distinct conversion behaviors
  - 11% Fast Adopters (Orange)
  - 14.3% Adopters (Red)
  - 58.8% Lazy Adopters (Blue)
  - 17.1% Churned Customer (Green)
- We also provide categorized buckets for the customers who only purchased the old products



## Sales Forecast

### Data

- Monthly average revenue data for 100 products from 2018Q1-2021Q5
  - We grouped non-core products by divisions to generate 27 product groups
  - We imputed the second and third quarters of 2020 to reduce covid effects



### Modeling

$$SARIMA(p, d, q)_{non-seasonal} (P, D, Q)_m_{seasonal}$$

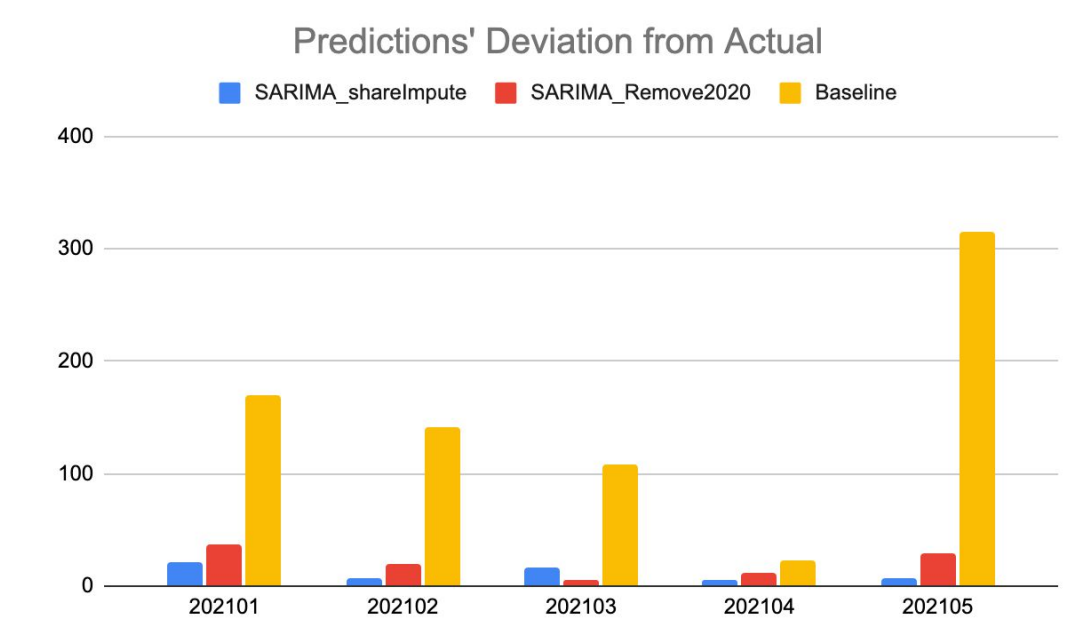
We used time series predictive model, SARIMA, for sales forecasting

Advantages of SARIMA model:

- Able to incorporate seasonality
- Interpretable parameters and results so that it's easy to incorporate with the domain knowledge from the sales team

### Results

- We improved the accuracy from the baseline prediction by 30%



## Challenges

### Missing Data



- The quarterly and monthly transplant data from UNOS public database
- The death rate and readmission rate for each transplant center in order to assess potential and optimize diagnostic portfolio
- Number of donors in each medical center
- Competitor's sales revenue and quantity information
- Current companies' Salesforce distribution
- Specific features of customers: operations or researches that they focus on

### COVID Effect



COVID has a strong impact on the revenue and product sales portfolio on mid 2020 and therefore the sales information on 2020 is not an accurate historic indication for future product potential

### Solution

- Eliminate the entire 2020's data and assume that 2021 happens right after 2019.
- Imputation**
  - Impute using SARIMA time series prediction
  - Impute using the monthly 1 on 1 ratio with the month in first quarter: assume that the first quarter 2020 is normal
  - Impute using the monthly percentage of each month

### Data Quantity



### Sparse data

- More than 90% of 0 value for some individual products sales revenue and some even include negative values

### Observation Match

- The customer names from the public dataset and the company's transactional dataset is different

### Limited Timespan

- Due to product update and new product release, we can only use sales revenue from 2018
- Solution:** Hardcore the parameter tuning range based on the seasonality pattern we observed in the aggregated level

## Business Impacts



- Detected estimated over 8.7M dollars of potential sales opportunities across 28 states in the U.S. for two core products
- Discovered and defined 4 distinct conversion behaviors and estimated the average conversion time of 3 years without promotion
- Estimated a customer churn rate of 17.1% and a conversion rate of 25.3% without promotion
- Improved the accuracy from the baseline prediction by 30% to help the management team better allocate salesforce

## Recommended Actions



- Synergize salesforce to the accelerated markets detected with abundant potential sales opportunities and high confidence
- Track current old-product-only customers and flag the customers once they start purchasing the new product
- Conduct Case Studies to specific customers in each bucket
- Experiment different promotion strategies for each cluster
- Update the resource data for the sales forecasting model frequently to obtain accurate product-level forecasting results and capture hidden pattern

## Reproducible Tools



- End-to-end machine learning pipeline that connects company's ETL database and generates sales forecasting for business intelligence dashboard with practical imputation for period affected by Covid-19

