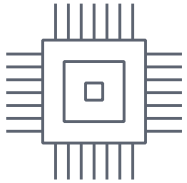


Revenue estimation for online marketplace products

Technology & SaaS business of law firm

Built prediction model to estimate the sales of products on online marketplace using azure Databricks.

Technology & SaaS business of law firm needs revenue estimation for online marketplace products

Picture this…

You're looking to build a prediction model for ecommerce sales by leveraging their ecommerce sales database for thousands of products across multiple categories and the corresponding product and seller attributes. You have utilized a third-party service to understand the estimated Amazon ecommerce sales for the products of interest of their prospective and current clients. The third-party sourced data is prone to data accuracy/ quality issues and has an associated recurring cost.

You turn to Accordion.

We partner with your team to build prediction model to estimate the sales of products on online marketplace using azure Databricks, including:

1) Increasing visibility into the revenue captured by unauthorized sellers, helped brands identify the dollar opportunity associated that can be recaptured by initiating/ setting up a legal framework for managing eCommerce channels.
2) Helping in prioritizing the products/ un-authorized sellers that should be targeted from legal standpoint and achieve quick wins.
3) Understanding the impact of holiday events on the product sales by the un-authorized sellers and take preventive actions.

Your value is enhanced.

- As the 3rd party relationship is essential for the client, leveraging the in-house prediction tool instead of the 3rd party paid service for 20% of the products helps client save ~USD 140K annually.
- Leveraged the model to predict product sales for prospective leads to identify the dollar opportunity accurately and improve the lead conversion rate.

REVENUE ESTIMATION FOR ONLINE MARKETPLACE PRODUCTS

KEY RESULT

- 20% of the products helps client save ~USD 140K annually.

VALUE LEVERS PULLED

Day level product sales estimation using product features based on regression models (xgboost)

E-commerce product sales estimation for a law firm

Revenue estimation for online marketplace products

Situation

- The client utilized a third-party service to understand the estimated Amazon ecommerce sales for the products of interest of their prospective and current clients. The third-party sourced data is prone to data accuracy/ quality issues and has an associated recurring cost.
- Partnered with the client to build a prediction model for ecommerce sales by leveraging their ecommerce sales database for thousands of products across multiple categories and the corresponding product and seller attributes.

Accordion Value Add

- Increased visibility into the revenue captured by unauthorized sellers, helped brands identify the dollar opportunity associated that can be recaptured by initiating/ setting up a legal framework for managing eCommerce channels.
- Helped prioritize the products/ un-authorized sellers that should be targeted from legal standpoint and achieve quick wins.
- Understood the impact of holiday events on the product sales by the un-authorized sellers and take preventive actions.

Impact

- As the 3rd party relationship is essential for the client, leveraging the in-house prediction tool instead of the 3rd party paid service for 20% of the products helps client save ~USD 140K annually.
- Leveraged the model to predict product sales for prospective leads to identify the dollar opportunity accurately and improve the lead conversion rate.

Methodology/Approach - Overall architecture (1/4)

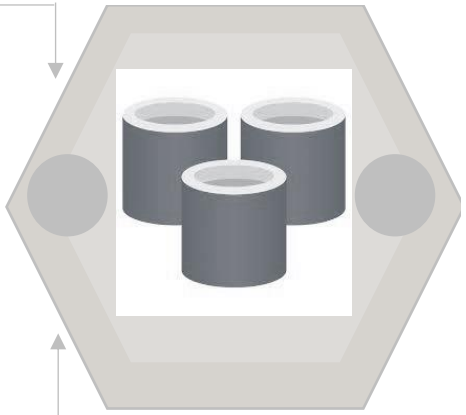
RPA

Daily Sales reports
- Seller level data

WEBSCRAPE

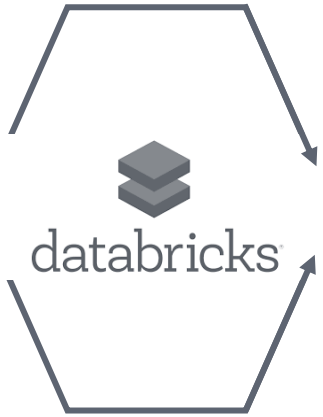
- 48 scrapes/product/day
- Product related Attributes
  - Rank
  - Ratings
  - # Sellers
  - # Variants

AGGEGATED DATA



- Seller level Product Sales is aggregated across all sellers & extrapolated to day level.
- Data from both the sources is aggregated and processed to extract relevant features.
  - Rank in category
  - Avg. List Price
  - Avg. Rating etc.

PREDICTION MODEL



Models are built at product category level to predict daily product sales based on its features.

OUTPUTS

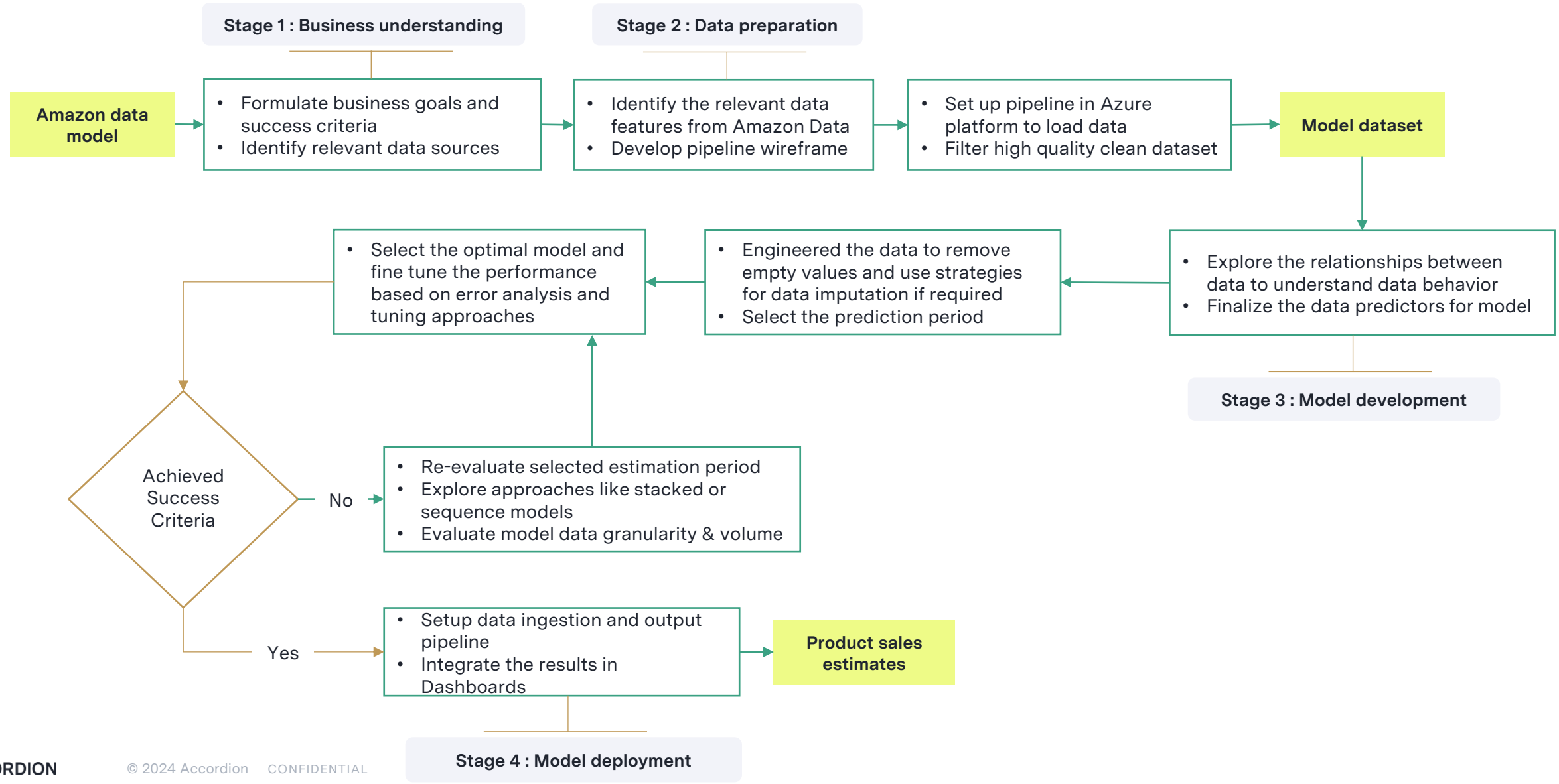
PERFORMANCE METRICS

- Mean Absolute % Error
- Median Absolute % Error

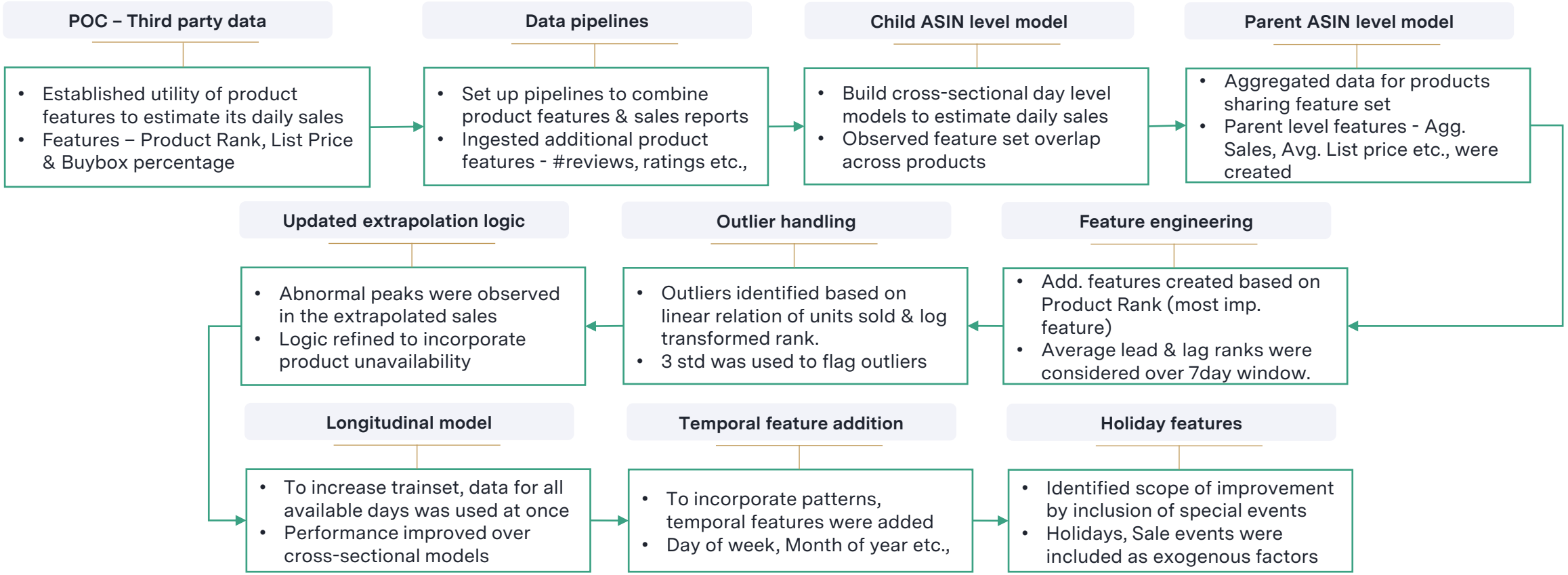
BENCHMARK COMPARISON

- Revenue Bucket wise
  - > $1000/day
  - $100/day - $1000/day
  - <$100/day

Methodology/Approach – Development process flow (2/4)



Methodology/Approach - Model iterations (3/4)



1. ASIN: Amazon Standard Identification Number - 10-digit alphanumeric product identifier code issued by Amazon to every unique item offered in its catalog and housed in its warehouses
2. Child ASIN: Represents a product available in amazon's catalog at the most granular level
3. Parent ASIN: Product representative of all the related variants that share feature set like ratings, reviews etc.,
4. Buy-Box: A section on the right side of an Amazon product detail page where customers can add a product to their cart directly and the sale is captured by the default seller.

Methodology/Approach – Model development milestones (4/4)

Child ASIN model: MAPE 254%

Model granularity - Day level individual models
Best model - Extra trees regressor
Features - List price, Rank in category, Number of reviews, Average rating, Number of variants, Product unavailability ratio

Parent ASIN model: MAPE 197%

Model granularity - Day level individual models
Best model - Random forest regressor
Approach - Aggregated data to parent hierarchy
Additional features - Avg. Category rank of product for preceding & succeeding 7 days.

Outliers' treatment: MAPE 163%

Model granularity - Day level individual models
Best model - Random Forest Regressor
Approach - Dropped the outliers based on log transformed Rank for the product for Parent ASIN level data.

Revenue extrapolation update: MAPE 63%

Best model - Random forest regressor
Approach - Updated the logic to incorporate product unavailability during the day.

Historical Data model: MAPE 56%

Model granularity - Single model with complete historical data
Best model – XG boost regressor
Approach - Train a single model with the historical data
Additional features - Holiday and discount days to incorporate days with larger sales for better learning in addition to existing features.

Comparison with third party service

The model has outperformed the third-party performance benchmark for the revenue segment of interest (>$1000/day)