Predicting Demand in Three Sided Marketplace

3/31/21 9:59 AM

Speaker: Alok Gupta (DoorDash)

Date: 03/31/2021

Marketplace Dynamics

Three sided market: consumers, merchants, dashers

Journey of an order

- Consumer opens App
 - Forecast web traffic
- Consumer sees restaurants with delivery time estimates
 - Predict delivery times
- Consumer selects a Merchant and creates an order and checks out
- Order gets offered to and accepted by the Merchant
- Order gets offered to and accepted by a Dasher
 - Assignment algorithm consumes forecasts for dashers
- Order is prepared by the Merchant
 - Predict food preparation time
- Order is picked up by Dasher
- Order is delivered to Consumer
- Dynamic delivery time prediction
- Consumer may contact DoorDash Support if there's an issue
 - Forecast number of support tickets

Different Demand Forecasts

Consumer Volume Hourly

- Hour by hour
 - For tracking and pacing
- 1 week out
 - To keep accuracy high
- Cut by submarket Very granular geographical cuts, e.g. up to zip code

Consumer Volume Daily

- Day to day
- 7 days out
- Cut by business line
- Manual inputs
 - o Requires back and forth with finance on endogenous factors (e.g. promotions) and exogenous factors (e.g. weather)
- **Enables balancing**
 - o Can do inter-day demand and supply shaping and create accountability around actions

Consumer Volume Weekly & Annual

- Week to week / Every quarter
- 6 weeks out / 4 quarters out
- Cut by business line & city / business line & country
- Manual inputs
- **Enables balancing**

Problem

Previous methodology used WoW from prior years to forecast

This requires lots of adjustment

Solution

Implemented a decomposed time series methodology

- First decomposes historical volume into current course and speed by removing the effects of key volume covariates
- Then applies a Holt-Winters exponential smoothing model to estimate current business trend and seasonality
- And finally layers back in the effects of key covariates

Model Architecture



Note the advantages Current course and speed volume effects can be separated from estimation of complex

- interacting patterns like weather or holidays This is helpful given the limited number of observations
- Results

- Saw a 13.6% relative volume weighted daily MAPE improvement This improvement is present even with known issues in weather, promotion, and monthly
- seasonality processing so we still have more room to improve

Problem

Annual Forecast

We have a number of levers we can pull to effect company metrics such as orders, revenue, profit, e.g.

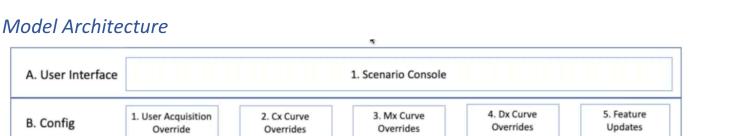
- Supply levels Demand levels
 - Fulfillment conditions
 - Marketing spend At annual planning level we would like to simulate different interventions to accurately

vertical, by use cohort)

- estimate the impact on company metrics to make better decisions
- Solution

We build an ML based self-service application that can simulate the long term impact of any strategic action on all of DoorDash's major metrics

- One-stop shop for sizing and costing future initiatives and drastically reduces the time it takes to make decision by consolidating different aspects of planning by different teams.
- Two advantages: Output is very rich as it allows for multiple cuts (e.g. geographically, by business
- The structure is very modular which allows for parallel iterations and performance check points



5. Daily Pacing 3. Mx Cohort C. Component 2. Cx Cohort 1. New Cx Forecast 4. Dx Cost Curves Models Curves Curves 2. Revenue 4. Profit 3. Cost 5. Traffic D. Summary 1. Volume Forecaster Forecaster Forecaster Forecaster Models Forecaster E. Output Layer 1. Data 2. Logging F. Visualization 1. Reports 2. Dashboards Layer

Organizing Demand Forecast Democratization / Platformization

Approach Start with an expert team of Forecasters

- on 2-3 most important demand Begin to generalize their methodologies
- into some common workflows / platforms
- Onboard other Data Scientists on how to use the platform for their use cases

Allows the expert team to test on

Advantages

- themselves first to create a high quality platform, and then scales them by creating a self-service platform. All future algorithm development on the
- platform will automatically lead uplifts in models using the platform
- Input data and pre/post-processing is consistent i.e. same corrections / anomalies handling