Customer Willingness to Wait for Specific SKUs

Aman Bhargava, Aditya Gupta, Simranjeet Khalsa, Arjun Dayal Mishra, Yang Wang

Krannert School of Management at Purdue University

Abstract

A model to predict the number of days a customer is willing to wait for a specific SKU based on historical data that will help companies can better anticipate customer needs and optimize the shelf space by stocking units for which willingness to wait is low. The model will eventually help in a company's aim to achieve SKU rationalization to avoid losing out on sales due to lack of inventory on hand. The willingness to wait for a SKU is dependent on multiple factors like part type, customer requirement and urgency, delivery time to the closest store, and availability of SKU with competitors in the area. Hence, the analysis needs to be done at a SKU level for a cluster of stores using survival analysis and regression techniques tracking the customer drop-off for products as delivery time increases. The support from client regarding providing of required data and legacy insights has been a key part of our project structure. Based on these findings, this research presents insights into predicting the time a customer is willing to wait for a specific SKU using time-series, survival and regression concepts in Python and R.

Keywords: Willingness to wait, Backorder acceptance, SKU, Python, R, Time-series, predictive modelling, delivery time, store inventory, average wait time, Survival analysis

Customer Willingness to Wait for Specific SKU

Brick-and-mortar stores of any retailer face the same challenge – optimizing the physical space available to best display and store products that customers from the area may want to buy. The store may not hold all the items the company has to offer to its customer base. Hence, companies need to decide on SKUs they want to hold in stores and hold the other SKUs in nearby stores or warehouses and have them delivered when required. However, some companies offer time sensitive products for which customers may not be able to wait for another store to return. Waiting for a product varies for each product depending on its requirement and the customer purchasing the product. For example, a customer may be willing to wait a few days for big ticket items like radiators or head gaskets while they may not be willing to wait for items like wipers or filters. Generally, a product is shipped from nearby stores or warehouses if a part requested by a customer is not at a store. The customer may or may not wait for the delivery time depending on several parameters like product type, customer requirement and urgency, delivery time for closest store, and availability of SKU with competitors in area.

We aim to predict the time the customer is willing to wait for each SKU based on the above parameters to better optimize shelf space and eventually minimize losses due to lack of product availability at the store. SKU rationalization can help a business make better business decisions, improve operations, and identify areas of improvement in the supply chain by reducing inventory carrying costs, less shrinkage due to stockouts or delivery time higher than acceptable.

The study's first objective is to identify if the SKU acceptable wait time for a customer is dependent on a store level for each SKU or the same across stores for each SKU. Next, acceptable wait time thresholds are identified for each SKU at the store level and the corresponding delivery times from nearby stores keeping the SKU. This would help us identify stores requiring an improvement in delivery times and shelf space optimization. The firm could identify the store's areas of improvement in supply chain and other operational efficiencies. Most significantly, the company can minimize losses through lost sales due to long wait time for a SKU by better stocking their shelves.

Tackling the above problems, we first need to identify the dependencies in the data and identify the important parameters and their corresponding weights to predict the acceptable wait time accurately. Also, we would have to identify if the dependencies are at a store level or are they independent of the stores by correlation techniques. Once the important variables are identified we need to build a model using regression techniques on Python utilizing concepts such as time-series modeling to predict the average wait times that a customer would be willing to wait.

The remainder contains a literature review on various criteria and methods used for estimating customer willingness to wait, followed by a section explaining the dataset used and a proposed methodology to solve the problems and questions discussed above. The paper then delves into further detail about the models used in the project, finishing with the results achieved via this project and the conclusions we can take from it.

Literature Review

Derhami, Montreuil, and Bau (2020) "Retail networks offering high-value products with a broad mix of models, such as cars, employ on-demand inventory transshipments to avoid lost sales due to incomplete product availability" (p.1). This aligns with our project as we are working on finding out the willingness of a customer to wait for a product that is not available at a retail store of a nationally acclaimed auto parts seller in the US.

Derhami, Montreuil, and Bau (2020) go on to say, "We assume that customers know their desired product before visiting a retail outlet (B&M stores) or are appropriately guided by sale representatives to find their desired product from the firm's product portfolio." (p.2). The following assumption is a part of our working scheme for our project, as well as we will be generating SKU wait times from the given dataset and matching them to their stores. Some of these stores also lie on the same supply routes for the client. This gives us an idea of how the supply chain affects the wait time that a store has for getting the required part in their inventory.

The above two variables thus define the inventory policy. Hariharan & Zipkin (1995) explained that the inventory policy changes from keep-in-stock to order when the demand lead time exceeds the supply lead time. If the retailer can match the delivery in timeliness to the demand lead time, they need not keep the SKU. This forms an important cornerstone for our project as well to estimate customer willingness to wait. This in turn can tie in to the client's SKU rationalization algorithms to help with better inventory stocking.

Alim & Beullens (2021) mentioned that the best solution proposed by other researchers to tackle delivery times is offering financial incentives like discounts by evaluating the trade-off the company receives from additional gains by lowering inventory or other system costs. They also state that customers may reject the wait time and expect to have products at their desired time even after these incentives. This would lead to a loss in potential sales as customers may prefer competitors in the same area. The proposed finding the best-combined inventory replenishment and offer synchronization strategy as a Markov Decision Process and solve it by backward induction to solve the problem. These principles could be implemented to solve rationalizations problems due to wait time rejections.

Hillard (2012) outlined the process to achieve SKU rationalization while maintaining sales volume and reducing complexity. He mentions the importance of selecting SKUs to be rationalized based on delivery times, delivery costs, and other things. These values are an important consideration in his model to calculate the sales impact and cost complexity analysis signifying the importance of wait times in a model for SKU rationalization.

Kurata (2014) depicts how to develop a model taking in Poisson demand and Bernoulli customer response to out-of-stock items as the two variables in terms of product availability effect on sales. It further investigates finding out the best inventory management system based on variables ranging from supply chain size to demand rate. These variables can be considered while building a predictor model for our project employing the techniques discussed.

Chen & Plambeck (2008) aptly stated, "If lost sales are not observed, the Bayesian optimal inventory level is larger than the myopic inventory level (one should "stock more" to learn about the demand distribution)." They discuss the implications when a product is not in inventory, stating that a customer may accept a substitute or not purchase from the store. They propose deriving the maximum likelihood estimators (MLEs) of the demand rate to find customers' probability of waiting. The study can be used as a basis for us to build upon by considering the above analysis and tying it back to the point made by Hariharan & Zipkin.

Title	Author	Summary
Customer-Order Information, Leadtimes, and Inventories	Hariharan & Zipkin (1995)	Inventory policy based on demand vs supply lead time
Improving inventory system performance by selective purchasing of buyers’ willingness to wait	Muzzafer Alim & Patrick Beullens (2021)	Markov Decision process by backward induction for inventory replenishment

Assessing product availability in omnichannel retail networks in the presence of on-demand inventory transshipment and product substitution	Shahab Derhami, Benoit Montreuil and Guilhem Bau (2020)	Employing on-demand inventory transshipments to avoid lost sales due to incomplete product availability
Achieving and Sustaining an Optimal Product Portfolio in the Healthcare Industry through SKU Rationalization, Complexity Costing, and Dashboards	David Hilliard (2012)	Importance of delivery time and tie-in to sales impact and cost complexity analysis
How does inventory pooling work when product availability influences customers' purchasing decisions?	Hisashi Kurata (2014)	Development of a model taking in two variables Poisson demand and Bernoulli customer response to out-of-stock items.
Dynamic Inventory Management with Learning About the Demand Distribution and Substitution Probability	Li Chen, Erica L. Plambeck (2008)	Deriving maximum likelihood estimators (MLEs) of the demand rate and probability that customers will wait for the product.

Data

The data used for this study has been provided by a major auto parts dealer based out of the United States. The data provided contains information regarding the SKU specifications, sales and store inventory for each store and SKU. The data also provides us with the information of delivery times between the 2 stores and the delivery time for a product sold. These can be used to estimate the days a customer is willing to wait. Also, the table provides information about the customer demographics for a cluster of stores in the same area. The detailed information about all data present can be seen in Table 1.

Methodology

The first step undertaken by the team was to clean the data received from the client and undertake data exploration to find key insights about the data. The data was cleaned, and missing values were imputed using methods suggested by the client. A list of SKUs that had been sold after being transferred was generated to use as a sample case for instances where the customer was willing to wait for a product. These cases where the customer was willing to wait were market as 1, signifying the occurrence of an event. Meanwhile, based upon the discussions with the client important variables were selected for basing the prediction. Finally, a survival analysis will be conducted on a SKU level using Kaplan-Meier plots and Cox proportional Hazard model. The team plans to study the drop-off rate of customers as time passes for a specific SKU. Lastly, the team explored various models to predict customer willingness to wait using SAS EM and selected the best model based on the MSE values.

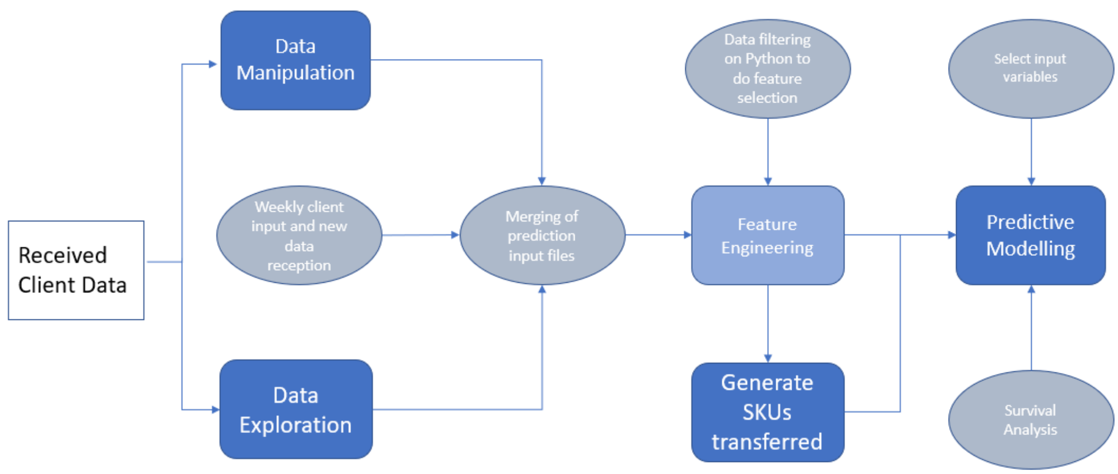

Figure 1. Methodology

Model

Survival analysis techniques such as Kaplan-Meier plots and Cox Proportional Hazards model were used to study each variable independently. Kaplan-Meier plots were used to study the customer drop-off rates by each categorical variables whereas a Cox Proportional Hazard models were used to identify the significance and impact of each variable.

Finally, a linear regression model was used on SAS EM to build a prediction model by stepwise selection of variables. The stepwise method would consider all combinations of variables and hence provides the best models. Other models like the Decision tree and HP Forrest were considered but later dropped due to poor performance.

Results

Firstly, a survival analysis was done to understand customer drop-off rates by different categorical variables to understand critical threshold values for the customer drop-off. This was achieved by constructing Kaplan-Meier plots in R like in figure 1 below.

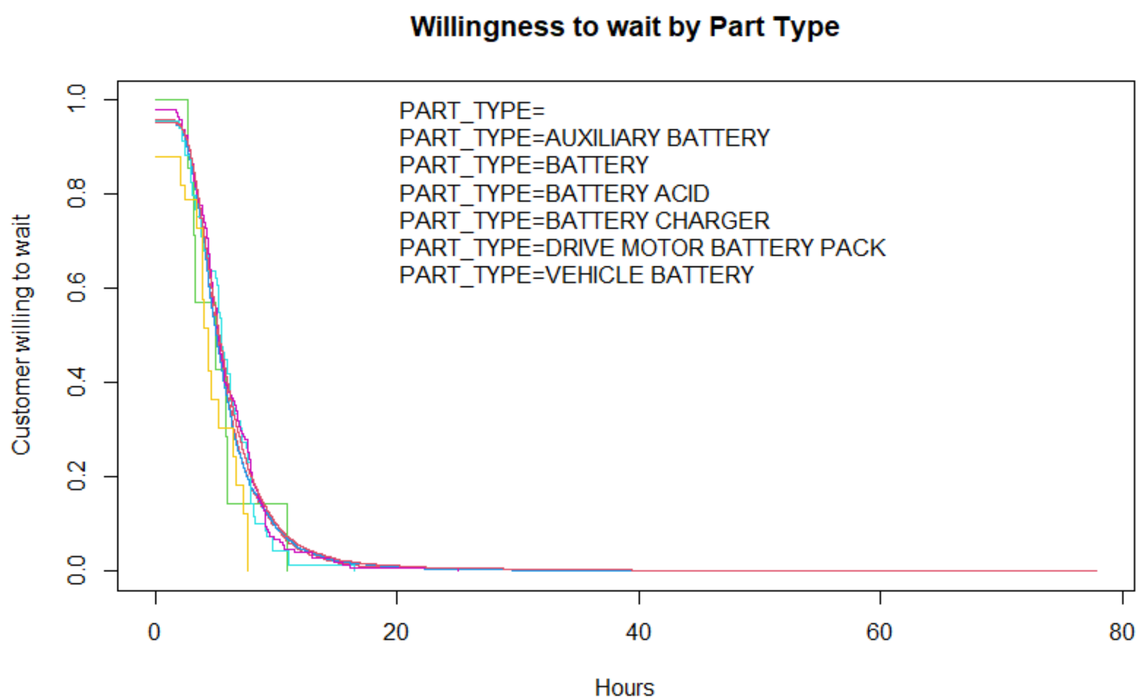

Figure 2. Customer willingness to wait rate by part type

Key insights were generated using Cox model, such as customers would be willing to wait 5% less for a vehicle as compared to another type as seen in figure 3. Similarly, plots were built using different variables to track corresponding drop-off rate and iterations of Cox models were run using bucketing on variables like age to explore trends and significance of different variables. Learnings from the survival analysis were then implemented into the different predictive models being built on SAS Enterprise Miner.

	coef	exp(coef)	se(coef)	z	p
STORE_NUMBER	8.797e-06	1.000e+00	1.756e-06	5.009	5.46e-07
PART_TYPEAUXILIARY BATTERY	3.930e-01	1.481e+00	3.784e-01	1.039	0.298954
PART_TYPEBATTERY	-5.176e-02	9.496e-01	2.107e-02	-2.456	$\boxed{0.014037}$
PART_TYPEBATTERY ACID	1.724e-02	1.017e+00	1.215e-01	0.142	0.887175
PART_TYPEBATTERY CHARGER	-2.226e-02	9.780e-01	8.405e-02	-0.265	0.791160
PART_TYPEDRIVE MOTOR BATTERY PACK	4.216e-01	1.524e+00	1.748e-01	2.411	0.015892
PART_TYPEVEHICLE BATTERY	-5.714e-02	9.445e-01	1.668e-02	-3.427	0.000611
POP_EST_CY	3.680e-05	1.000e+00	7.517e-07	48.953	< 2e-16
POP_DENSITY_CY	-5.070e+01	9.531e-23	1.085e+00	-46.744	< 2e-16
PCT_WHITE	-6.681e-01	5.127e-01	2.637e-02	-25.339	< 2e-16
AGE	-1.008e-02	9.900e-01	2.295e-03	-4.393	1.12e-05
PCT_COLLEGE	6.072e+00	4.337e+02	1.691e-01	35.913	< 2e-16
PCT_BLUE_COLLAR	-6.836e-03	9.932e-01	8.027e-03	-0.852	0.394391

Figure 3. Significant predictors as per Cox model

Upon running various models in SAS EM (Linear regression, Decision Tree, Gradient Bossting, HP Forest, etc) we found the linear regression model to be the best as per the selection criteria of mean squared error. The significant variables identified using stepwise selection included product lifecycle, sales in the geographical area, periods the SKU was in stock, delivery time, median household income in area, part type, population density, SKU number itself and cost of SKU.

Conclusions

Estimating the time a customer would be willing to wait for a SKU is of utmost importance and would eventually help the company with SKU rationalizing, optimizing their shelf space. Thus, in this study we created outputs that could be used to identify the SKUs for which a customer is not willing to wait or in other cases estimate the ideal number of days a customer would wait for a particular SKU. This would then feed in as an input for the client to feed into their SKU Rationalization algorithms, eventually helping them minimize losses due to lack of SKU in inventory when required.

The customer's willingness to wait can be estimated using a regression model on some of the important variables mentioned in the results. Additionally, the firm could use survival analysis techniques to further study the different drop-off rates as per the categorical variables and identify the underlying importance of each variable and the impact it has on wait times.

The business could see many benefits from the above study such as reduced inventory cost, better prediction to avoid lost sales, knowledge of important features to track for better prediction, reducing wait time by optimization, and importantly improve customer satisfaction.

References

Derhami, S., Montreuil, B. & Bau, G. (2020). Assessing product availability in omnichannel retail networks in the presence of on-demand inventory transshipment and product substitution. Omega *102*(4). DOI:10.1016/j.omega.2020.102315

Hariharan, R., & Zipkin, P. (1995). Customer-Order Information, Leadtimes, and Inventories. Management Science, *41*(10), 1599–1607. http://www.jstor.org/stable/2632740

Alim, M. & Beullens, P. (2011). Improving Inventory System Performance by Selective Purchasing of Buyers' Willingness to Wait. *European Journal of Operational Research*, https://www.sciencedirect.com/science/article/pii/S0377221721006226

Hilliard, D. (2012). Achieving and Sustaining an Optimal Product Portfolio in the Healthcare Industry through SKU Rationalization, Complexity Costing, and Dashboards. Massachusetts Institute of Technology. http://hdl.handle.net/1721.1/73385

Kurata, H. (2014). How Does Inventory Pooling Work When Product Availability Influences Customers' Purchasing Decisions?. *International Journal of Production Research*, *52*(22), 6739-6759, DOI: 10.1080/00207543.2014.916825

Chen, L. & Plambeck, E. L. (2008). Dynamic Inventory Management with Learning About the Demand Distribution and Substitution Probability. *Manufacturing & Service Operations Management, 10*(2), 236-256. https://doi.org/10.1287/msom.1070.0165

Tables

Table 1

Data Dictionary

Table	Column	Description
tops_down_skus	sku_number	unique number used to internally track a inventory
tops_down_skus	merchandise_group_desc	product category/ bpg
tops_down_skus	qty_sold_ppy	quantity of product sold past past year
tops_down_skus	qty_sold_py	quantity of product sold past year
tops_down_skus	qty_sold_cy	quantity of product sold per current year
tops_down_skus	sum_py_qty_sold_on_hand	sum of product stocked in the store past year
tops_down_skus	sum_cy_qty_sold_on_hand	sum of product stocked in the store current year
tops_down_skus	sum_cy_qty_sold_transfer	sum of product transfered from other stores current year
tops_down_skus	sum_ppy_qty_sold_on_hand	sum of product past past year in the store
tops_down_skus	sum_py_qty_sold_transfer	sum of product transfered from other stores past year
tops_down_skus	sum_ppy_qty_sold_transfer	sum of product transfered from other stores past past year
tops_down_skus	num_stores_unit_sales_oh_gt0	unit sales for number of stores greater than 0
tops_down_skus	lookup_cnt_cy	lookup count current year (looked up product for customer but didn't sell; demand quantity)
tops_down_skus	lookup_cnt	lookup count
tops_down_skus	failure_sales	related to vio
tops_down_skus	cy_ts_forecast	time series forecast
tops_down_skus	lost_qty_cy	lost sales current year
tops_down_skus	ss_sales_cy	specific sales type
tops_down_skus	sales_signal_cy	sales indicator for current year (may be similar to trend)
tops_down_skus	cy_unit_sales	current year sales
tops_down_skus	projected_growth_pct_cy	projected growth percent
tops_down_skus	other_unit_pls_lost_sales	subsection of lost sales
tops_down_skus	other_unit_pls_lost_sales_cy	subsection of lost sales for current year
tops_down_skus	weighted_lookup_cnt_cy	weighted looked count current year
tops_down_skus	weighted_lookup_cnt	weighted lookup count
tops_down_skus	cy_periods_in_stock	current year periods in stock (13 periods, 4 weeks each)
tops_down_skus	cy_sales_cost	current year sales cost
tops_down_skus	pop_est	population estimate
tops_down_skus	lifecycle	useful life of part
tops_down_skus	sku_existence_cy	sku existence current year
tops_down_skus	ppy_unit_sales	past past year unit sales
tops_down_skus	age	age demographics
tops_down_skus	pop_density	population density
tops_down_skus	ss_sales	subsection of sales
tops_down_skus	other_unit_pls_lost_sales_py	subsection lost sales past year

tops_down_skus	total_vio_cy	total vehicales in operations current year (based on store area)
tops_down_skus	median_household_income	household income
tops_down_skus	other_gross_sales	gross sales
tops_down_skus	adjusted_lifecycle	useful life of part (might be adjusted on surrounding area)
tops_down_skus	sales_signal	sales signal
tops_down_skus	lifecycle_pre_peak_post	pre if new sku, peak if prime selling years, post for older cars/past prime
tops_down_skus	total_vio	total vehicles in operation
tops_down_skus	adjusted_lifecycle_cy	adjusted useful life
tops_down_skus	unit_sales	unit sales
tops_down_skus	sold_since_maxi	sold since added to store
tops_down_skus	part_type	part type
tops_down_skus	sku_store_pdq	sku store (may be a 0/1)(Is 0 or 12)
tops_down_skus	py_sales_cost	past year sales cost
tops_down_skus	ppy_sales_cost	past past year sales cost
tops_down_skus	pct_blue_collar	percent blue collar
tops_down_skus	lifecycle_cy	lifecycle of product
tops_down_skus	unadjusted_total_vio	unadjusted total vehicles in operation
tops_down_skus	sku_existence_py	sku existence vehicles in operation
tops_down_skus	pct_college	percent college
tops_down_skus	py_gross_sales	past year gross sales
tops_down_skus	application_count	sku-specific types of vehicles it would fit
tops_down_skus	pct_of_lifecycle_remaining	percent of lifecycle remaining
tops_down_skus	cy_gross_sales	gross sales
tops_down_skus	sku_store_pdq_cy	sku store pdq current year
tops_down_skus	projected_growth_pct	projected growth percentage
tops_down_skus	establishments	census data, registered businesses in area
tops_down_skus	road_quality_index	road quality index
tops_down_skus	mpog_id	part of sku hierarchy / category
tops_down_skus	unadjusted_total_vio_cy	unadjusted total vehicles in operation for current year
tops_down_skus	other_unit_sales	other unit sales
tops_down_skus	pct_white	census level, percent white people
tops_down_skus	other_sales_cost	other sales cost
tops_down_skus	ppy_gross_sales	past past year gross sales
tops_down_skus	wait_time_bucket	takes into account some people waiting for product
tops_down_skus	lost_qty	lost sales current year
tops_down_skus	qty_sold_cy1	quantity sold current year
tops_down_skus	qty_sold_py1	quantity sald past year
tops_down_skus	qty_sold_ppy1	quantity sold past past year
tops_down_skus	trend	qty_sold_py-qty_sold_ppy
tops_down_skus	cat	sales category bins (0-2,26-100,100+)
tops_down_skus	cat_py	category past year (should be the same overall)

tops_down_skus	filter_reason	filter reason (superceded, discontinued, etc.)
bottoms_up_gt/le	pop_est	population estimate
bottoms_up_gt/le	pop_density	population density
bottoms_up_gt/le	sku_store_pdq	sku store pdq
bottoms_up_gt/le	total_vio	total vehicles in operation
bottoms_up_gt/le	lost_qty	lost quantity
bottoms_up_gt/le	ss_sales	sales
bottoms_up_gt/le	avg_cluster_unit_sales	avg unit sales for cluster
bottoms_up_gt/le	avg_cluster_lost_sales	average lost sales per cluster
bottoms_up_gt/le	adjusted_avg_cluster_sales	adjusted average cluster sales
bottoms_up_gt/le	qty_sold_py	quatity sold past year
bottoms_up_gt/le	qty_sold	quantity of product sold
bottoms_up_gt/le	avg_cluster_total_sales	avg total sales for cluster
bottoms_up_gt/le	lifecycle	lifecycle of product
bottoms_up_gt/le	adjusted_lifecycle	adjusted lifecycle of product
bottoms_up_gt/le	adj_avg_cluster_lost_sales	adj avg lost sales
bottoms_up_gt/le	adj_avg_cluster_total_sales	adj avg total sales for cluster
bottoms_up_gt/le	unit_sales	total unit sales
bottoms_up_gt/le	projected_growth_pct	projected growth percent
bottoms_up_gt/le	qty_wt0_ppy	wait time 0 past past year
bottoms_up_gt/le	qty_wt0_py	wait time 0 past year
bottoms_up_gt/le	bpg	base product group
bottoms_up_gt/le	store_number	store id
bottoms_up_gt/le	sku_number	sku id
bottoms_up_gt/le	part_type	what type of part
bottoms_up_gt/le	platform_cluster_name	average for cluster for platform
bottoms_up_gt/le	pct_white	percent of population that is white
bottoms_up_gt/le	age	age
bottoms_up_gt/le	pct_college	percent of population that is in college
bottoms_up_gt/le	pct_blue_collar	percent that is blue collar
bottoms_up_gt/le	median_household_income	median household income
bottoms_up_gt/le	road_quality_index	quality of road
bottoms_up_gt/le	pct_of_lifecycle_remaining	Percent of lifecycle remaining
bottoms_up_gt/le	qty_sold_cy	quantity sold current year
bottoms_up_gt/le	filter_reason	special reason type
bottoms_up_gt/le	trend	trend year over year
results	store_number	store id
results	sku_number	sku id
results	prediction	prediction from model
results	wait_alpha	wait time coefficients for expo function
results	wait_beta	wait time coefficients for expo function