Numbersbright

EBOOK ILLUSTRATION

A GUIDE TO NUMBERSBRIGHT DATA ANALYTICS

Feh Gonne

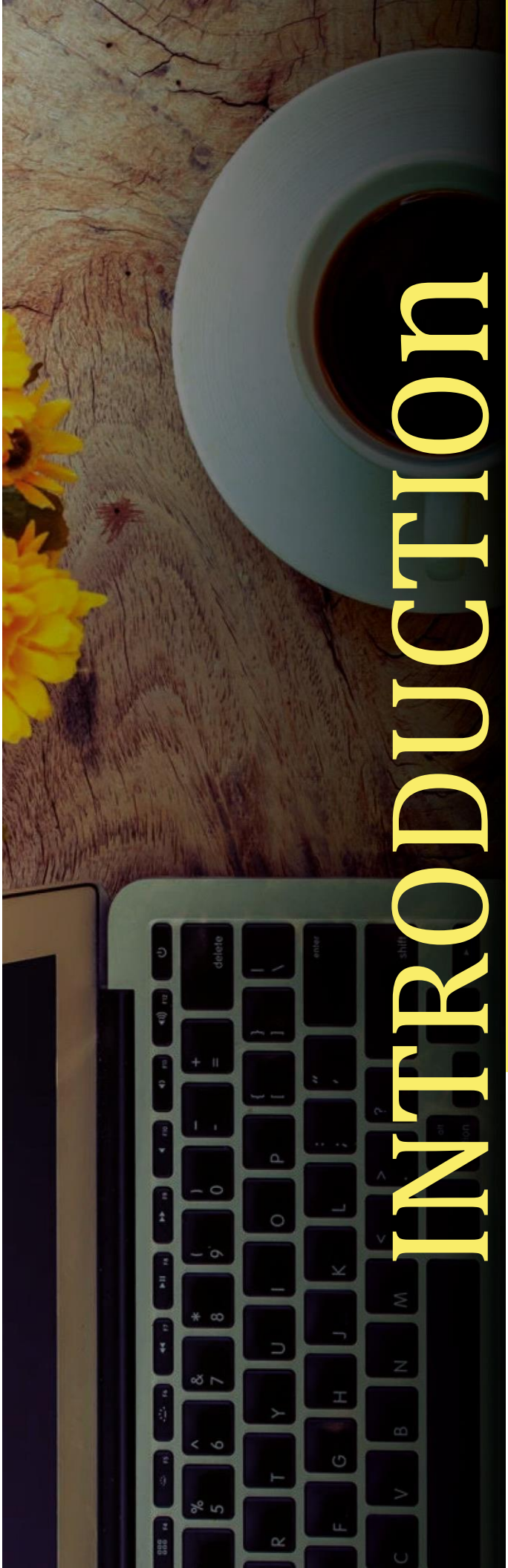
EBOOK ILLUSTRATION

A GUIDE TO NUMBERSBRIGHT

DATA ANALYTICS

Feh Gonne

TABLE OF Contents



INTRODUCTION

The future belongs to data analytics, machine learning, and AI (artificial intelligence). The global economy depends significantly on how data is collected and used to infer valuable information to make well-formed decisions. Only those businesses will be victorious in the next Industrial Revolution who will successfully decipher the art and science of using their business data to hone their product strategies. Data Analytics effectively improves digital reach for businesses by helping them personalize the customer experience and optimize their decision-making capabilities. At Numbersbright, data analytics is the magic that turns your data into business intelligence and derives opportunity and value from your growing data. Numbersbright gives the best data analytics services to the wholesale, retail, and services sector. It is a leading business analytics company that assists you in:

» Leveraging Deep Learning

» Data Visualizations

» AI (Artificial Intelligence)

» Machine Learning.

We use your business data to create value for your business by implementing an integrative approach towards commissioning business intelligence projects.

Numbersbright Ebook - A Guide to Numbersbright Data Analytics In order to learn ow the Numbersbright data analytics process, you need to download and study the Numbersbright Ebook, which is a comprehensive guide for every business to understand how your business can benefit from our sought-after services. If you are looking to identify the correct customer segmentation, this Ebook can help you understand how we do this at Numbersbright LLC. The Numbersbright Ebook will help you identify the right product for your business and boost your revenue using sought-after technologies such as Matlab, Python, R, Power BI, Tableau, High charts, D3.js, and Micro Strategy. The Numbersbright Ebook will also guide you on how you can increase your brand loyalty and articulate the hidden value of your services and products with reduced market risks, increased throughput, and more thoughtful decisions.

At Numbersbright, we enable you to monetize data for your business, and the Numbersbright Ebook will enlighten you on how.

GETTING THE DATA READY

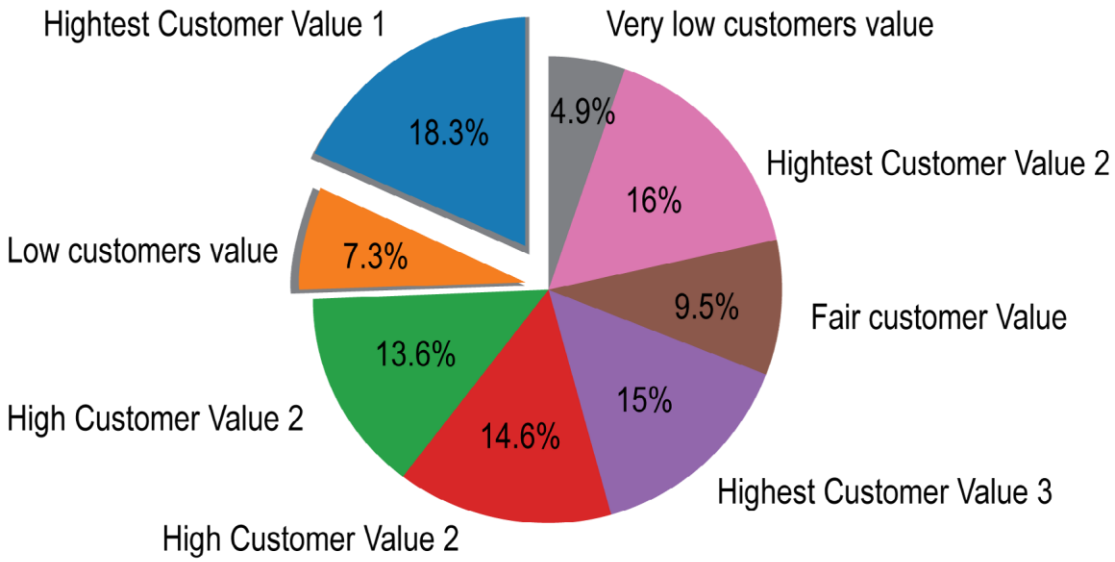
(DATA PREPROCESSING)

At Numbersbright, the first step to data analytics is preprocessing the raw data in order to lessen how much excess informational indexes.

FEATURES

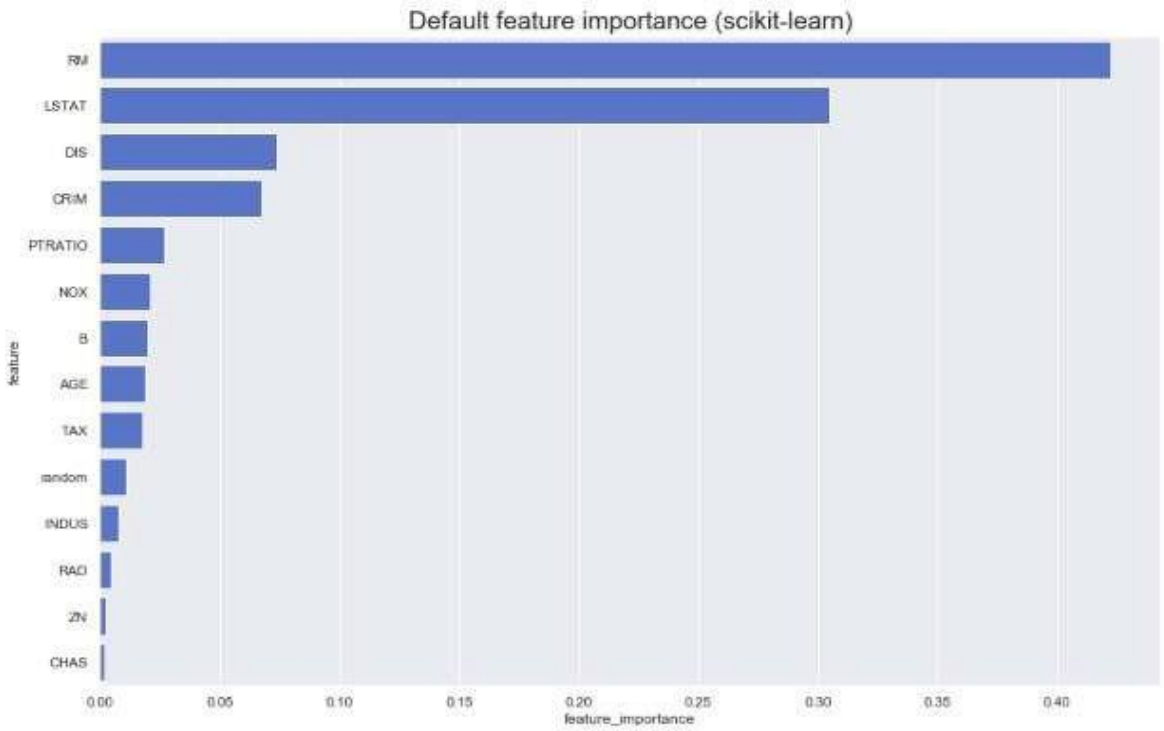
EXTRACTION

We use the Feature Extraction technique to extract new features, which are a linear combination of the existing features. This technique is the beginning of an innovative and valuable data analytics approach. It helps decrease the number of assets required for handling without losing important data.



At Numbersbright, we measure the significance (as example: highest, high, fair,and low value) of future by ascertaining the expansion in the model's expectation error in the wake of permuting the feature.

FEATURES IMPORTANCE

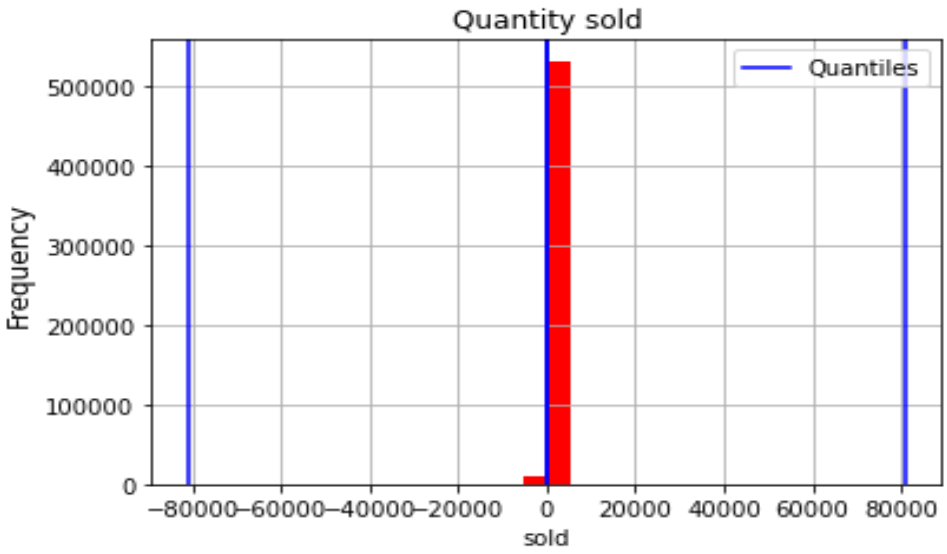


After Feature Extraction, we move on to Feature Importance, the technique that computes a score for every information high- lighted in a given model. The scores represent the "importance" of each feature extracted through the Feature Extraction technique; a higher score(0.40 =RM as above) suggests that the particular feature will significantly a fact the model utilized to foresee a specific variable.

At Numbersbright, we use an advanced future predictive model to analyze the most critical features selected. It helps us focus on things that matter the most for your business's success.

Next step, it is to find out what values set itself out of data analysis perceptive.

OUTLIER' VALUES



An outlier is an unusual observation that lies an abnormal distance from the mass of data. **Numbersbright** data analytics investigate this data that lies outside the other values**(-80000,+80000**) in the set to conclude what will be thought of as strange.

Also, some data contains **missing values** leading outcomes analysis to the wrong path.

IMPUTATIONS OF MISSING VALUES

At **Numbersbright**, the next step in data preprocessing is the implementation of the standard methodology of Imputation. This methodology functions admirably for the most part. The attributed values may be methodically underneath or above their genuine qualities, but we use the best models and best approach algorithms to make better predictions and fulfill the missing values by thinking about which values were initially absent. In addition, it allows an extra lift by following what values had been credited and tracking what values had been imputed.

PRODUCT

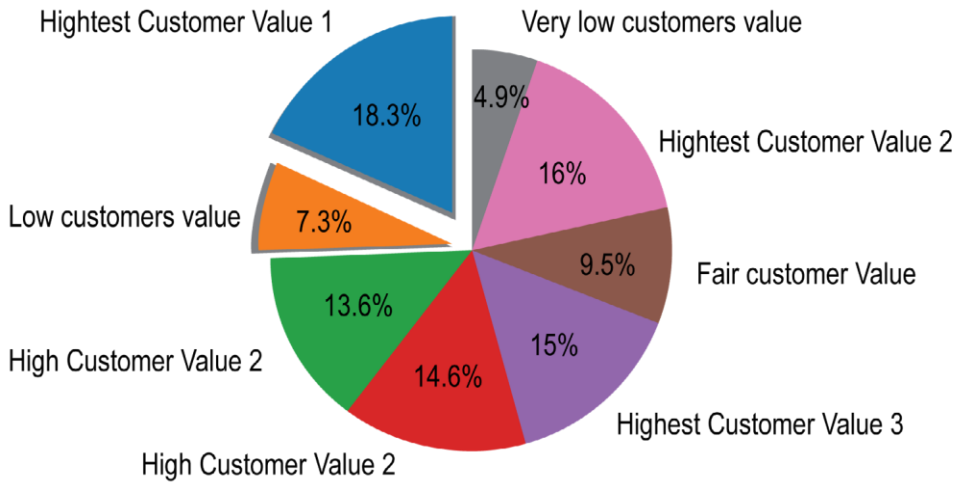
SEGMENTATION

IDENTIFYING

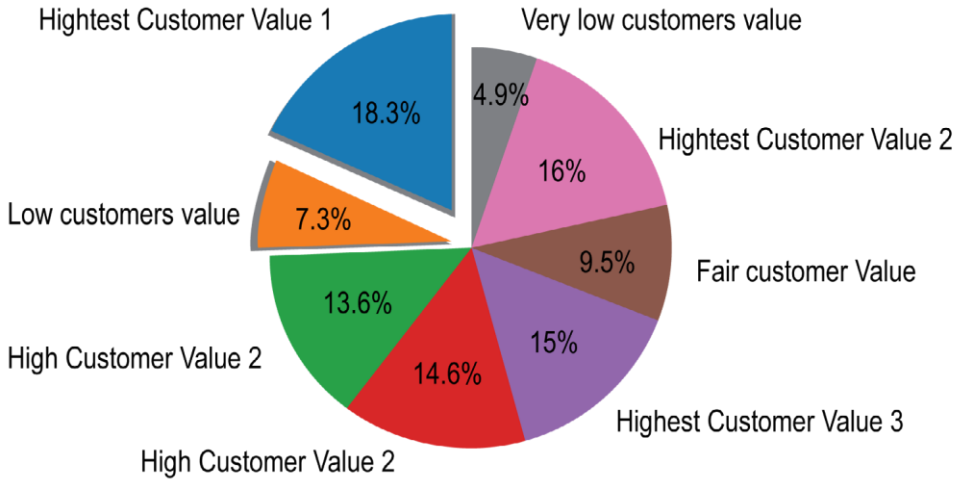
PRODUCT SEGMENT
BASED ON CUSTOMER PURCHASES

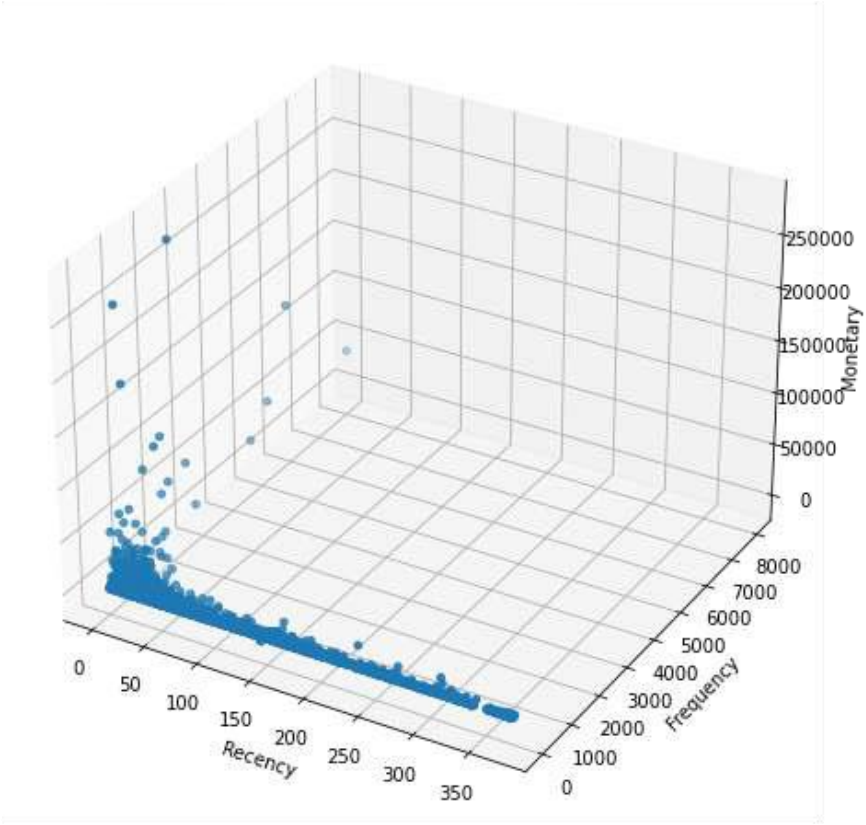
Product identification is the key to developing valuable business strategies. In most cases, an increase in product purchase value would significantly increase customer value.

PRODUCT PURCHASED VALUE



CUSTOMER VALUE





At Numbersbright, we begin by identifying the customer value by calculating:

» **Recency:** What did you buy and when did you buy?

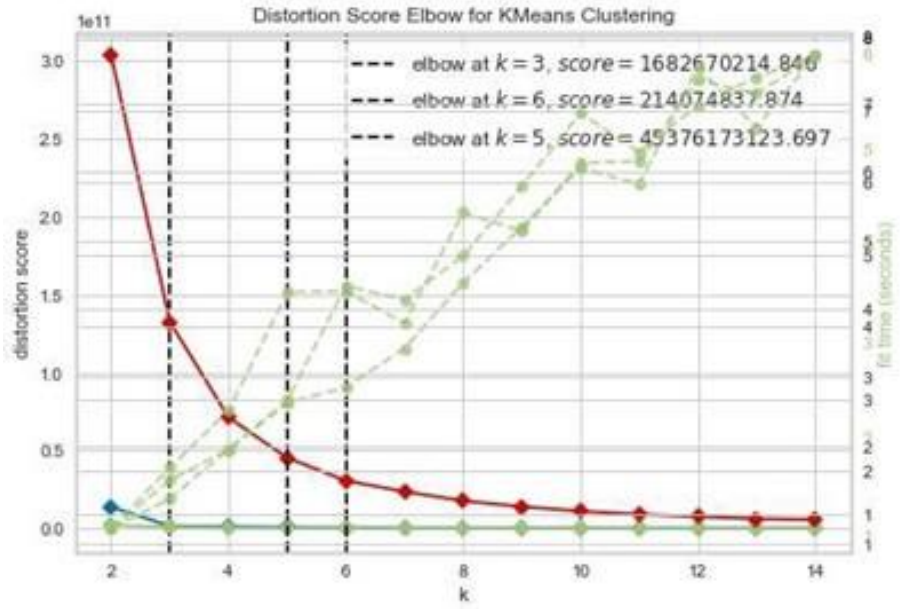
» **Frequency:** How often did you buy this product?

» **Monetary:** How much do you spend on each purchase?

As mentioned earlier, three fundamental questions are essential to set up a customer segment, but they alone do not provide enough information to develop the customer segment adequately.

DETERMINING CLUSTERS

SUFFICIENTLY RELIABLE FOR EACH CASE

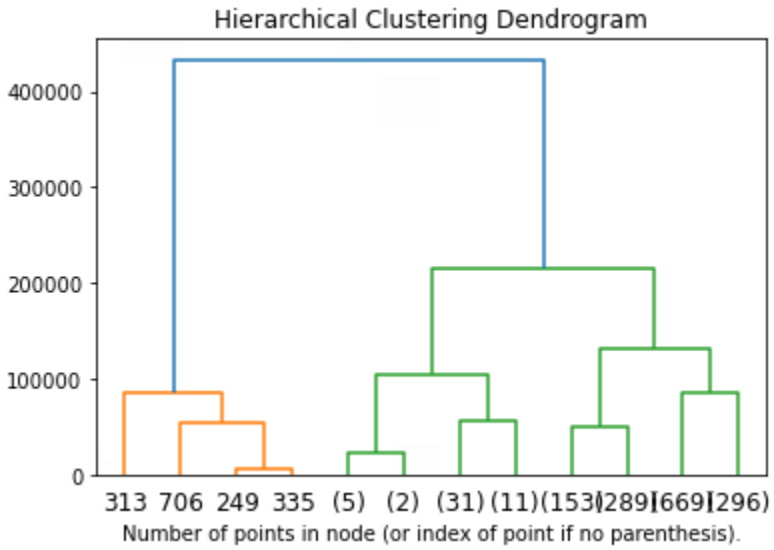


To develop the product segment, at Numbersbright, we use reliable algorithms that compute the similarity between an object in one cluster contrasting with another cluster. For instance, there are three features show by vertical discontinued points in the graph above. Hence, when k=3, it means 3 clusters. In other words, there are the most important subgroup into this feature that Numbersbright will use for any product segmentation.

HIERARCHICAL RELATIONSHIP AND FINDING

PRODUCTS SEGMENTS

At Numbersbright, we go further than just customer segment and product segment identification but the most profound understanding of the products or services offered by wholesale, retail, and services sectors.



Even after determining clusters and identifying several clusters coming from different features like products, prices, or invoices ID, we still don't know how sub-clusters are organized inside each feature. For this, we at **Numbersbright** utilize a robust algorithm that clearly shows segments inside each cluster feature and illustrates the relationship between different clusters like customer behavior and psychographic as shown in this graph above.

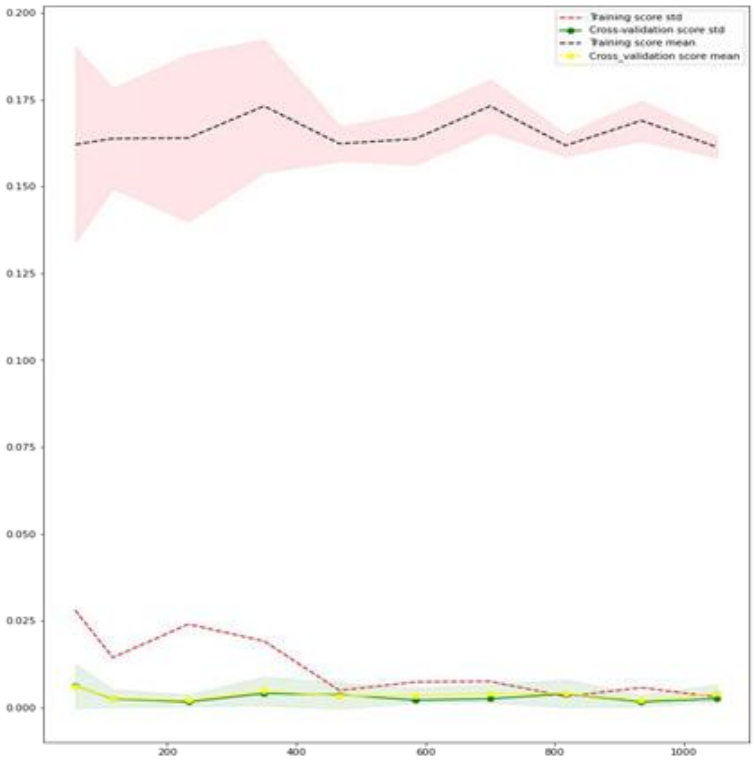
CUSTOMERS

SEGMENTATION

INTRODUCING RANDOMIZATION And

variance under-control

At this stage of customer segmentation, there is a large variance between low and high frontiers shown by “training score mean” and “cross-validation score mean,” making it hard to target with precision the segmentation model.

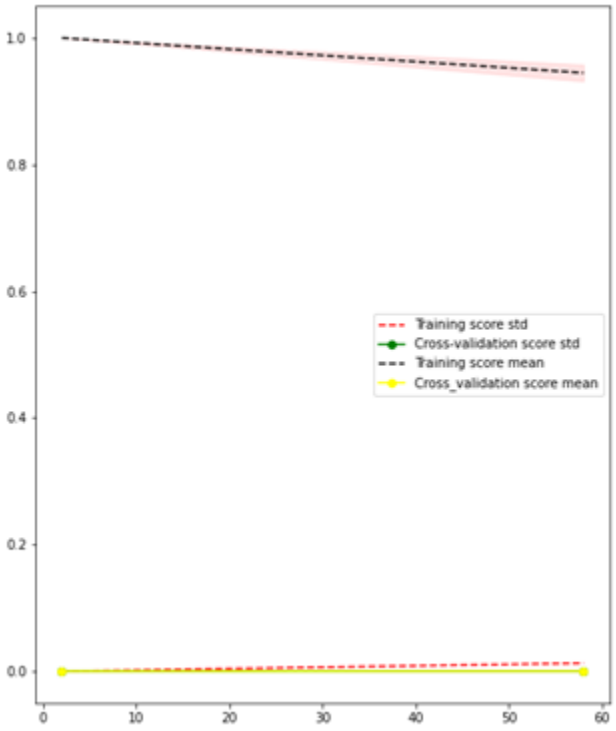


At Numbersbright, we continue to investigate our model and change our model until we are happy with the approval score we get. We apply the randomization technique at this stage to control the lurking variable, establish a cause and effect relationship, and ensure that the results are accurate. At this stage, our model is yet to call the last model.

AUGMENTED PREDICTIVE ACCURACY

Of customer segmentation

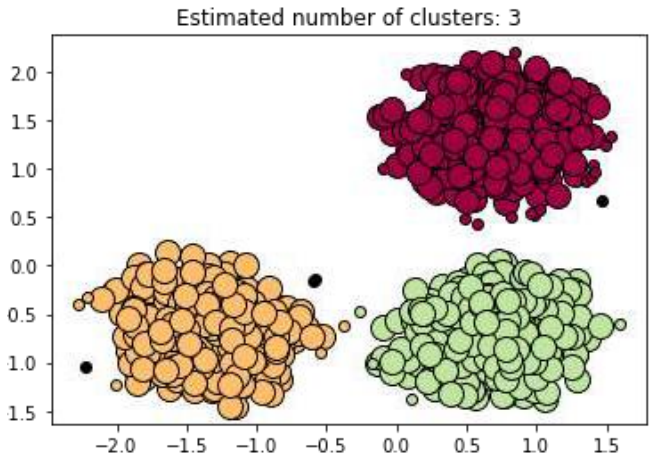
At **Numbersbright** we use the right algorithm to minimize errors by combining many new learning models or previously used models to build robust predictive customer segmentation.



We achieve the most accurate segmentation within the shortest lower and higher frontiers by dividing the customer's behavioral and psychographic as accurately as possible.

IDENTIFYING
THE BEST CUSTOMER'S SEGMENTATION

- Estimated number of clusters: 3
- Estimated number of noise points: 4
- Homogeneity: 1.00
- Completeness: 0.98
- V-measure: 0.99
- Adjusted Rand Index: 1.00
- Adjusted Mutual Information: 0.99
- Silhouette Coefficient: 0.69



Here come the final result, we identify inside each features the number of clusters clearly. For instance, K=3, depicts three different clusters show in orange, red and green color (reduce from fat frontier to 3 errors in black spot) and as above.

At **Numbersbright,** we improve inaccurate algorithm to the most accurate within less errors possibilities and achieve the best customer segmentation.

REVENUE GROWTH

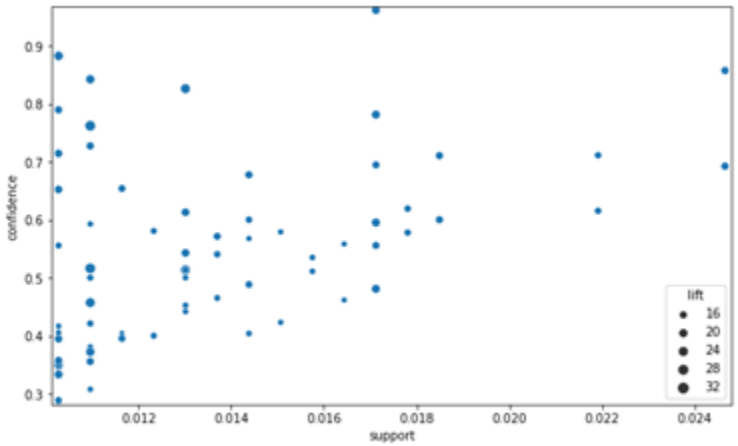
FULL ANALYSIS

(HIGHLY RECOMMENDED)

BASKET ANALYSIS

OR PRODUCTS AFFINITY

At **Numbersbright**, we carry out Market Basket Analysis to recognize items that clients need to buy. We empower deals and business marketing teams to foster more viable item situations, evaluating, strategically pitch, and up-sell methodologies using the market basket analysis.

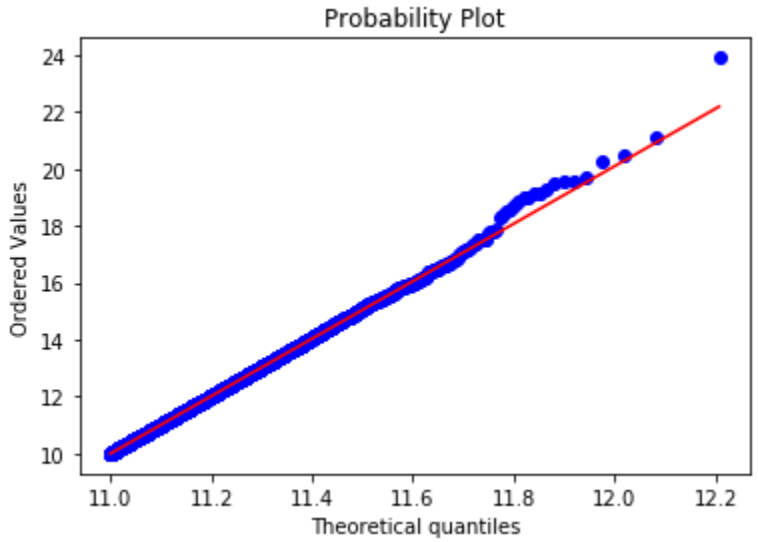


We use reliable and accurate algorithms which help us identify the customer shopping cart of goods purchased. We analyze large amounts of transaction history and recognize which items customers often buy together based on their purchasing behaviors. This data analysis results help the business better address the product needed.

DETERMINING

THE PARETO LAW OF PRODUCTS AFFINITY

We at **Numbersbright**, use the Pareto principle, also known as the 80/20 rule to help businesses shift their endeavors towards practices that genuinely make a difference to their customers and primary concern.



We broadly concentrate on ways of behaving and examples of customers to decide ties in buys with the goal that we can take advantage of them to increase the cross-selling potential of our client's businesses.

DETERMINING
PRICE COMPONENTS

Any graph variation gives valuable data about item, products, deals and sales around then. For our examination or study concerns, weighting past and present information is the beginning to understand the impact on services, production, sales and purchases.

At Numbersbright, we determine the business cycle based on the product, price, and customer interactions to better understand the price component. We do the time series examination to define the price point of the product or service.

» Seasonal variations

» Cyclical fluctuations

» Secular trend

» Irregular variations

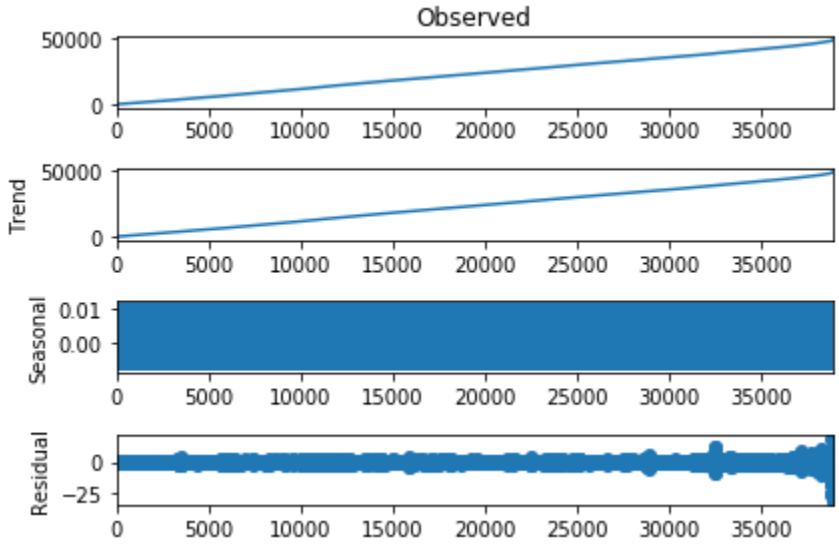
In addition, to learning about all the above techniques and methodologies, Numbersbright's EBook also discusses errors and interpretations as follow:

ERRORS

INTERPRETATIONS

The primary objective of advertising or marketing in this sense is to get your products and services before customers who will find them beneficial . . . in any case, what's the most ideal way to do that? The last thing that your customers need is to be targeted with messages and advertisements that don't connect with them and that is the place where Numbersbright comes in with data behavior analysis. After weighting data behavior, we can easily forecast production, sales or purchases in the future approach. In other words, companies can predict sales and revenue in the future or take some precautions if needed.

FINDING INTERNAL FLUCTUATIONS OF PRICE AND NOISE COMPONENT



We realized that the costs of services and products have the capacity of organizing or deciding financial plans concocted by individual subjects or economic activities. There is a mechanism of price analysis that helps to determine internal fluctuation. When this mechanism is able to function satisfactorily, price changes will invigorate or control the market interest. At Numbersbright we determine your business cycle base on your product, price and customer interactions to better understand the price component and how does is behaved through this cycle.

We can consider prices as times series because they are matched all the characteristic such as

» Secular trend, which describe the movement along the term;

» Seasonal variations, which represent seasonal changes;

» Cyclical fluctuations, which correspond to periodical but not seasonal variations;

» Irregular variations, which are other non-random sources of variations of series;

In marketing calculations and econometrics, and specifically in time series examination, an ARIMA auto-regressive integrated moving average model is a speculation of an auto- regressive moving average model, “integrated” being the difference here. Both of these models are fitted to time series information either to more readily get the information or to anticipate future focuses in the series. We do ARIMA to study the price behavior along the time and to detect all errors that can be minimized for better forecasting.

Dep. Variable: SalePrice No. Observations: 1460 Model: SARI- MAX(0, 1, 0)x(1, 1, [1], 12) Log Likelihood -18951.745

Date: Tue, 22 Sep 2020 AIC 37909.489

Time: 12:49:07 BIC 37925.321 Sample: 0 HQIC 37915.398 - 1460 Covariance Type: opg

---

coef std err z P>|z| [0.025 0.975]----------

---------- ar.S.L12 0.0218 0.037 0.586 0.558 -0.051 0.095 ma.S.L12 -0.9815

0.018 -55.582 0.000 -1.016 -0.947 sigma2 1.865e+10 1.89e-13 9.86e+22 0.000 1.86e+10 1.86e+10

Ljung-Box (Q): 447.96 Jarque-Bera (JB): 552.17 Prob(Q): 0.00 Prob(JB): 0.00 Heteroskedasticity (H): 0.94 Skew: -0.02 Prob(H) (two-sided): 0.51 Kurtosis: 6.03

Using python and machine learning we extract all errors and to be minimized to zero as shown below:

Errors Data

0 208500.000000

1 -27000.000000

2 42000.000000

3 -83500.000000

4 110000.000000 ...

1455 19.878494

1456 27504.152049

1457 48734.371245

1458 -119817.549130

1459 5976.130962

SETTING UP THE

NULL HYPOTHESIS

AND TESTING IT

AUGMENTED DICK-FULLER TEST

Augmented Dickey Fuller test (ADF Test) is a common statistical test used to test whether a given Time series is stationary or not. It is one of the most commonly used statistical test when it comes to analyzing the stationary of a series.

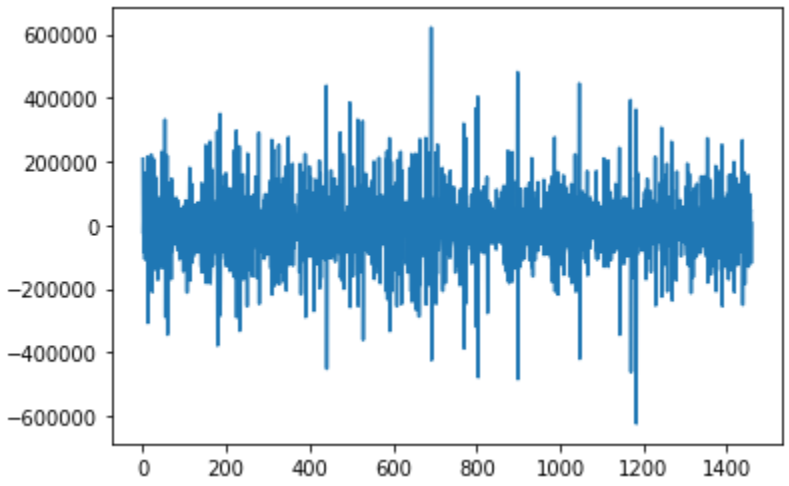
The primary differentiator between the two tests is that the ADF is utilized for a larger and more complicated set of time series models. The augmented Dickey-Fuller statistic used in the ADF test is a negative number. The more negative it is, the stronger the rejection of the hypothesis there is a unit root.

The null hypothesis of DF test is that there is a unit root in an AR model, which implies that the data series is not stationary. The alternative hypothesis is generally stationary or trend stationary but can be different depending on the version of the test is being used

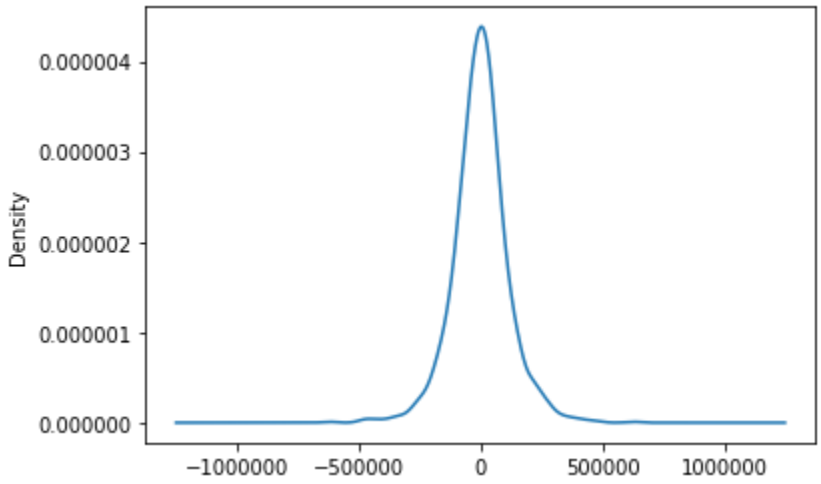
If there are unit roots, the series is not stationary. Accordingly, if the p-value of z(t) is not significant, the series is not stationary. If $z \leq z0.05$ then we reject the null hypothesis H0 that the series has a unit root. If there are no unit roots, then we conclude the series is stationary.

At Numbersbright, we go further than giving a test results, we give you the deepest meaning of it. Here, your data has no unit root and it is stationary because the shift in time did not change the distribution shape. For instance, the variation of your product's prices does not have impact on customer's behavior consumption. In other words, despite that the price has increased customer will still buying your product.

ERRORS GRAPH



ERRORS DISTRIBUTION GRAPH



We minimize errors as closer as to zero when the highest density is close to zero.

Understanding errors will help decide the over-fit element. As such, it shows anomalies point that don't match the model or the analysis. Fortunately, there are less errors within distribution around zero point Hence the prediction here is better. In addition to these, we also have explicit standards for how to reliably communicate the vulnerability related with data, it is better to know that any prediction can come without errors if any, they are hexogen errors source.

As shown above, this errors distribution graph around zero and the predicted analysis can be considered as a good one.

SETTING UP THE NULL HYPOTHESIS AND TESTING IT

Augmented Dick-Fuller test

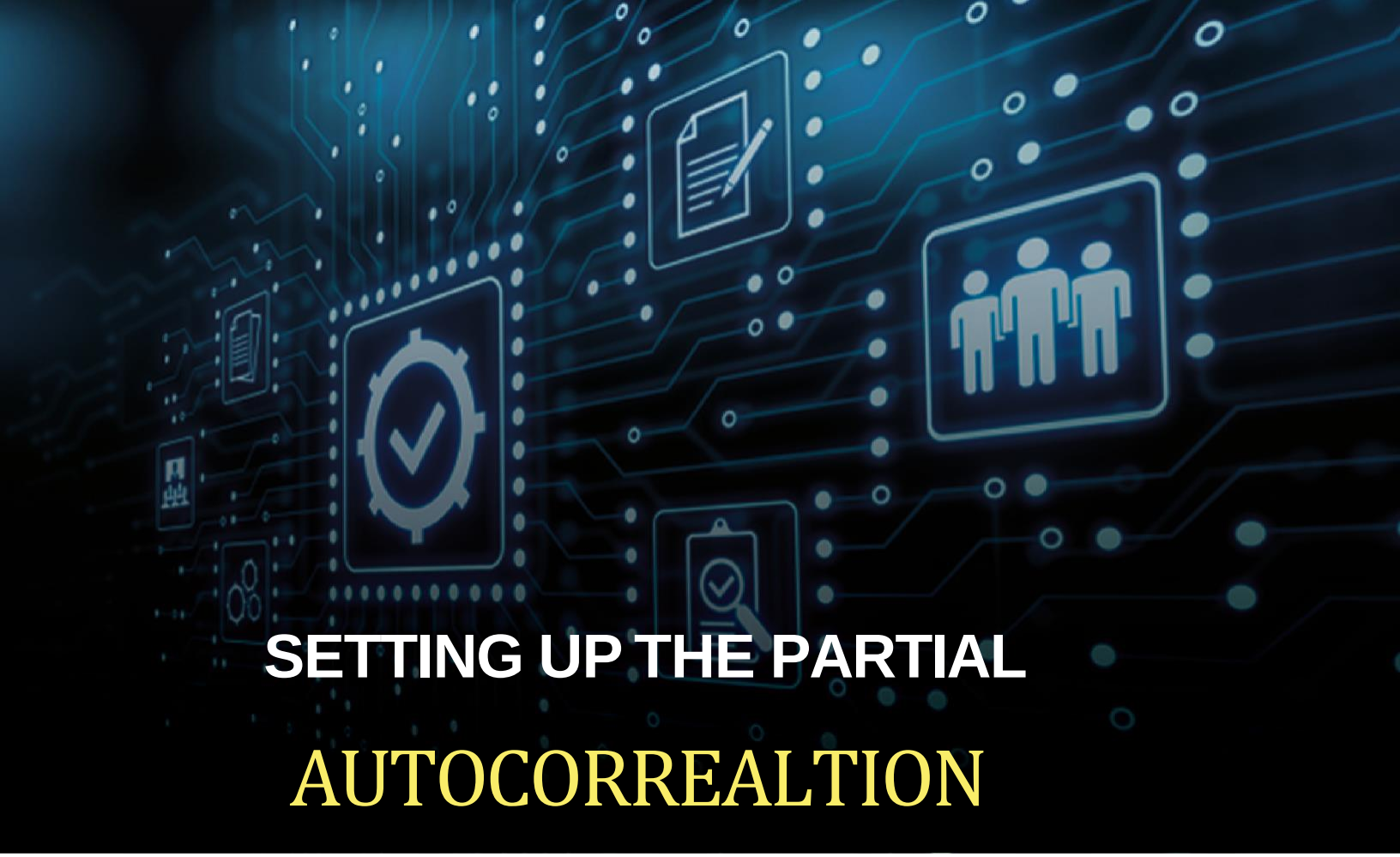
ADF test statistic:-6.652036048982912

P-value: 5.092258278926617e-09

# of lags: 21

Number of observations used: 1438

Strong evidence against null hypothesis: Reject null hypothesis Data has no unit root and it is stationary.



SETTING UP THE PARTIAL AUTOCORREALTION

THE AUTOCORRELATION

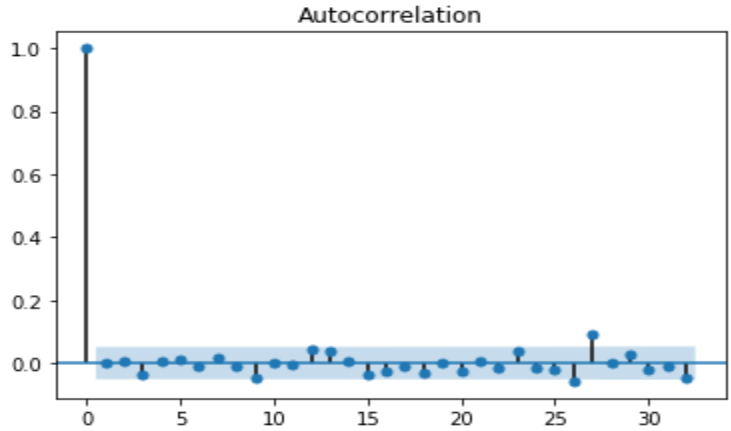
Using the partial autocorrelation and autocorrelation functions together to identify ARIMA models to be used, we look out for the following patterns on the partial autocorrelation function, we then go ahead to examine the spikes at each lag to determine whether they are significance. We know that a spike that's close to zero is evidence against autocorrelation. Therefore a significant spike will extend beyond the significant limits, which indicates that the correlation for that lag doesn't equal zero.

Large spike visible at lag 1 is followed by a damped wave that switches back and forth among positive and negative relation- ships. A higher order moving in normal term in the information is also seen. We utilize the autocorrelation capacity to decide the request for the moving normal term.

WHY AUTOCORRELATION IS IMPORTANT?

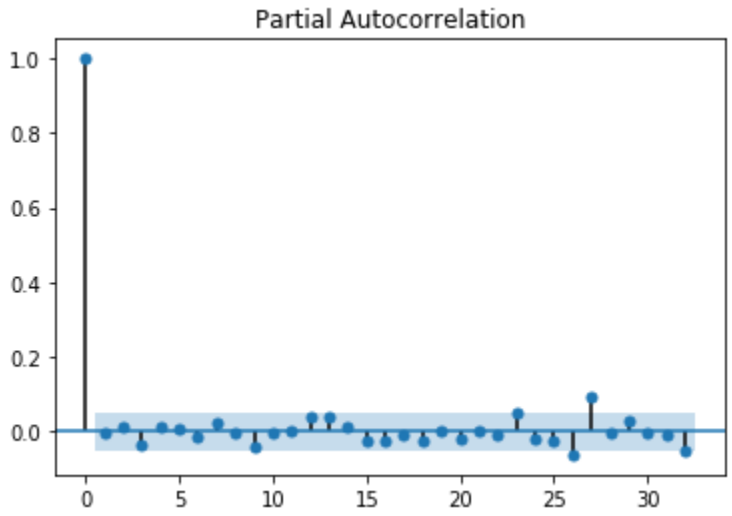
We know that the ACF describes the autocorrelation between observation and another observation at a prior time step, including direct and indirect dependence information. As needs be, the ACF is an element of the deferral or lag which decides the time shift taken into the past to appraise the likeness between main informative elements. At Numbersbright, we use the best algorithm to sort out your data potential when the previous model is incorrectly specified we can help you uncover patterns in your data, successfully select the best prediction model, and accurately assess the adequacy of the model to take care of issues in conventional analysis (OLS). Analyzing the autocorrelation function (ACF) and partial autocorrelation function (PACF) in conjunction is necessary for selecting the appropriate ARIMA model for any time series prediction.

"Weak stationary", meaning no systematic change in the mean, variance, and no systematic fluctuation. We also say weak stationary occurs at the point when the



time-series has steady mean and difference over the course of the time, put simply we mean that there is no trend. When performing ACF it is fitting to eliminate any pattern present in the information analyzed and to ensure the information is fixed. This algorithm helps us to find out how your revenue is perhaps affected long in time, and recognize issues that can be addressed to keep your business development top notch.

THE PARTIAL AUTOCORRELATION



PACF or partial auto-correlation function, would give the incomplete correlation of fixed time series with its own lagged values. It is also a partial auto-correlation function. Basically instead of finding correlations of present with lags like ACF, it finds correlation of the residuals (which remains after removing the effects which are already explained by the earlier lag(s)) with the next lag value, that is, regressed the values of the time series at every smaller or shorter lags. Hence 'partial' and not 'complete' as we remove already found variations before we find the next correlation. With this method we select and keep only the relevant and significant features for better correlations.

In some cases as above partial autocorrelation and autocorrelation show the same variation means principals features that impact your business revenue is stationary. In other words, it might mean that your customers behavior, psychographic and an affinity product did not change, thus, **you save money because any new marketing campaign is needed.**

FINDING OVER COST

AND MONEY WASTING

OVER COST

Over cost means the extra costs currently to some extent caused despite everything to be brought about in a specific year, normally including the changes from the two earlier years. We find the over cost by determining the forecast price of services or products and compare to the actual price. Only three outcomes can be found from estimation and calculation.

First, the forecast price of products or services and the actual price is zero (FP-AP=0). This result shows that prices will not change for the forecast price, exception of hexogen source of events as such war or an accident.

Second, the forecast price of products or services and the actual price is greater than zero (FP-AP>0). This result shows that prices will change and be higher than the current price for the forecast price. This difference of price might find explanations from inflation and taxes regulations.

Third, the forecast price of products or services and the actual price is less than zero (FP- AP<0). This situation is rarely observed in economy and demands more attention than those situation above.

At Numbersbright, we deeply analyze any situation, case by case and we always find viable substitute solutions that match your business goals.

MONEY WASTING

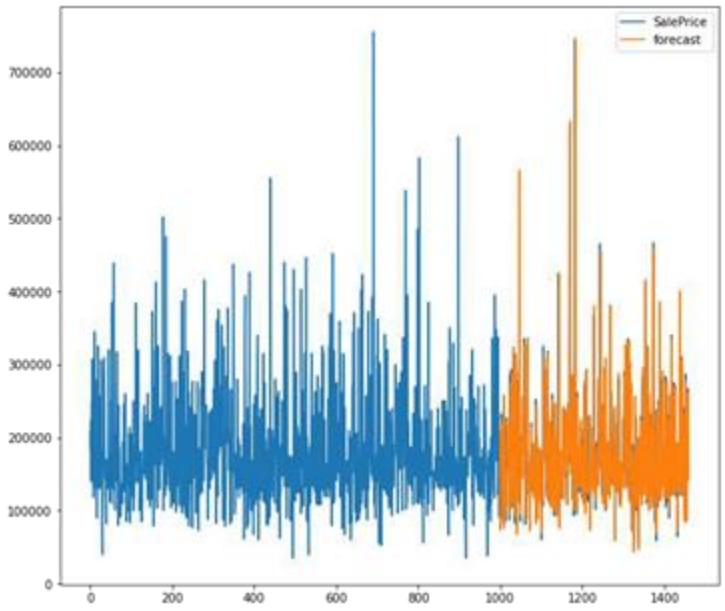
We at Numbersbright would rather not see cash gets squandered into promoting. Particularly when you're an independent venture, and you don't make as much as you spend on marketing. We are all about how your business uses their revenue efficiently according to the over-cost analysis. We calculate the wasting money in percentage when (FP-AP)>0 and when FP-AP<0.

In both cases FP maybe greater or less than AP and the results are FP>AP or FP<AP respectively and we find out when and where the wasting money efficiently equal to (FP-AP)/AP= x%

FORECASTING PRODUCTS OR SERVICES PRICE

AND ANTICIPATIVE ACTION

The objective of Forecast-based Financing of products and services is to lessen the effect of calamities. With the guarantee of expanded viability, the concept and techniques employed by **Numbersbright** is rapidly gaining impressions in analyzing data and predicting customer's trend. Clearly, our methods can be a useful tool to assist ventures, companies and organizations who might some way or another experience a calamity.



MONTE CARLO

SIMULATION OF PURCHASE IN STORE

We can make build some explanations as follow:

1. **They may not found the product they were looking for in store**

How do you explain that?

Do you at any point notice customers behavior in store when they simply circumvent the store without purchasing any items?

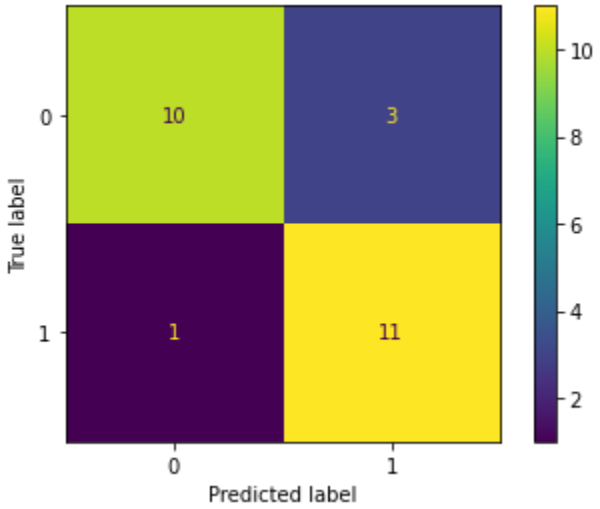
2. **Wrong products location or products characteristics**

   mis-understanding from both customer and store management.

In both situations (1&2) the store is losing money, at Numbersbright we have a powerful algorithm that uses Monte Carlo simulation to solve this problem and avoid losing money. This model helps us anticipate the likelihood of various results when the mediation of irregular factors is available. Monte Carlo reproductions help to make sense of the effect of chance and vulnerability in expectation and forecasting models. We find the customer high probability to not miss out your promotional or lower purchase rate products by increasing their choice of action.

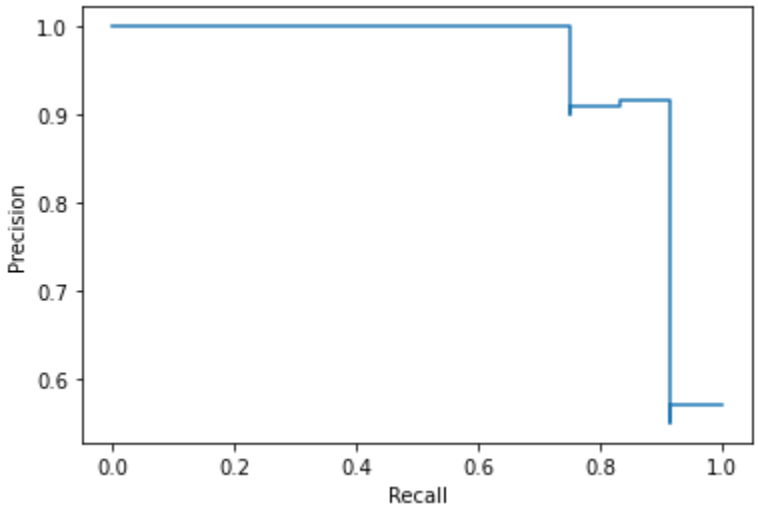
PRECISION AND ACCURACY
ALGORITHM TESTED

Confusion Matrix



In confusion matrix the quantity of right and inaccurate forecasts are summed up with count values and separated by each class. At **Numbersbright** we test all classification models for their predictive analytics and accuracy to recognize mistakes that would cynically or negatively be able to affect the outcome.

PRECISION

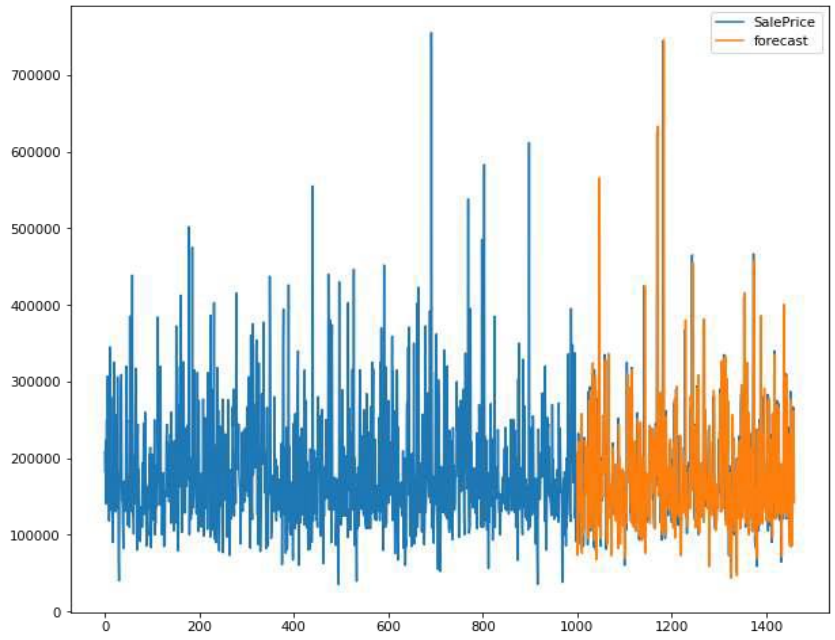


Precision is concerned with your model's predictions of positive examples. The value embodied by precision is especially clear in contemporary goods and service marketing. Customers hope to limit costs while purchasing. Precision showcasing would not just diminish the purchasing cost and power of customers but also cause an increment in the absolute worth of customers. In modern e-commerce the organization is customer driven. How- ever, precision measures something specific, precision is interested in the number of genuinely positive examples your model identified against all the examples it labeled such product segmentation, customer segmentation and price prediction.

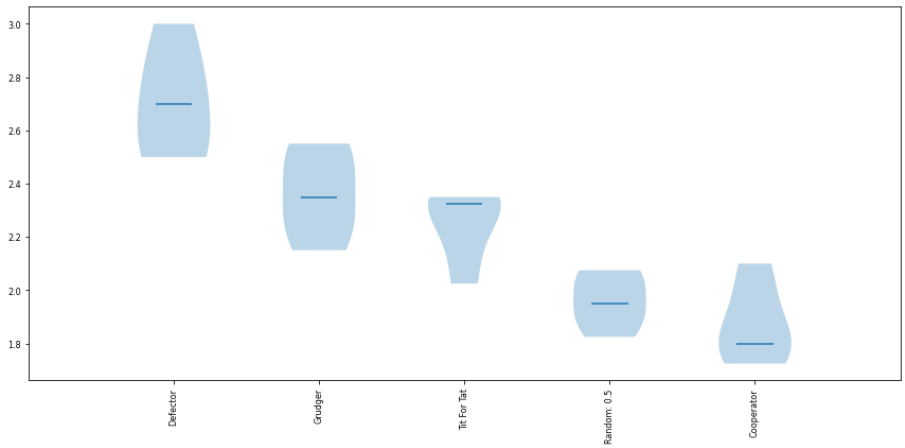
ACCURACY

Most promoting information is somewhere in the range of 10% and 20% in accuracy. Information that is under 10% accuracy for the most part doesn't cut it. Information that is better compared to 20% precise has diminishing crowds to sell. For each rate point more exact than 20%, the income for the marketing information merchant declines by 1%.

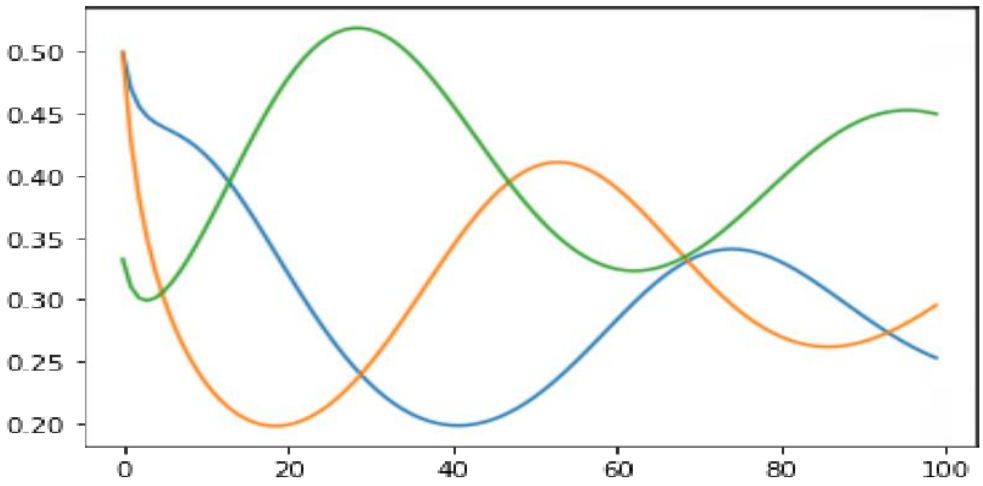
Accuracy defenses your total number of true predictions in total dataset that describes the percentage of the right precision such as the precision of the sales price forecast as shown below:



At **Numbersbright**, any product or service price designate to the market we help your business make the right decisions against competitor's strategy by increasing model precision. However, without knowing what the other will choose, it is possible each competitor will choose in such a manner that they both end up at an off-diagonal outcome that is outcomes that are not part of the main or initially proposed outcome. This strategy, give the maximum payoff.



Calculation of Pareto Optimal solutions incurs a lesser computational cost than that for Nash Equilibrium solutions.



At Numbersbright we find the business ideal that gives more benefit (green convex line at 0.50) than your competition, and the target price that cannot be exerted under the equilibrium (graphs intersection at 0.32).

CONCLUSION

The **Numbersbright Ebook** guides you through the process of Numbersbright data analytics using machine learning and AI (artificial intelligence). The Numbersbright Ebook gives better insight into data analytics, leading to an enhanced and guided decision-making process and improved performance. Numbersbright gives the best data analytics services for enhanced operational efficiency. We are a team of professional data scientists having certification in Business Analytics ,Financial Accounting and SQL for finance (Master of Financial Engineering program) respectively from Harvard Business School (Online) and Baruch College. We assist in business analytics that unlocks multi-fold outcomes for wholesale, retail, and services businesses.

Numbersbright data analytics services help reduce data handling complexities and allow companies to focus on operational efficiency better. The Ebook explains how utilizing the power of all disruptive next-generation data technologies helps in more precise customer segmentation and product segmentation. At Numbersbright, we implement the most advanced data technologies like Al

(Artificial Intelligence) and Machine Learning to enhance customer satisfaction by manifolds and offer a significant competitive advantage. If you want to generate new revenue and profit streams, you must focus on innovation based on business data insights and Guess Right and Brighter with Numbersbright.

Mailing address: 600 Broadway, Ste 200-3680, Albany ,12207 NY

Physical location: 14 Wall Street , Manhattan ,10005 NY

Customers support : support@numbersbright.com

Phone contact: 8778960032| +1 2126181234