If you are doing business and you are not into TV Advertising, you are losing business.

Author - Sonali Sheth – 29/01/2022

# If you are doing business and you are not into TV Advertising, you are losing business. It is the most dependable and traditional way of advertising. TV advertising in siti channel appeals to masses at large.



|  |
| --- |
| INTRODUCTION: |
| If you're selling a product or service and want to succeed right away, you need to communicate it to people through advertising and marketing.  Television, radio and newspapers are the most important media for effective offline marketing. However, over the last 15 years, these top three marketing tools have changed significantly.  This article describes some of the key changes that have occurred and how they can have the greatest impact on your advertising budget. |
| *“I know half my advertising budget is being wasted, if only I knew which half”* |
| PROBLEM DEFINITION: |
| The advertising dataset captures sales revenue generated with respect to advertisement spends across multiple channels like radio, tv and newspaper. The distribution strategy and the channel design have to be right the first time. It is not possible for clients to directly increase sales of the product, but they can control the advertising expenditure in each of the three media.  Therefore, if we determine that there is an association between advertising and sales,  then we can instruct our client to adjust advertising budgets, thereby indirectly  increasing sales.  The case study of Sales channel includes the detailed study of TV, radio and newspaper channel and predict the total sales generated from all the sales channel. |
| ABOUT THE DATSET: |
| Features:   * TV: advertising dollars spent on TV for a single product in a given market (in thousands of dollars) * Radio: advertising dollars spent on Radio * Newspaper: advertising dollars spent on Newspaper   Target:   * Sales budget in thousands of dollars |
|  |
| ***Observation:***   * **By, reviewing Data Set at initial stage, found continous datatypes; Target variable data is sales which seems to represent regression problem.** * **Each market is an observation, but each column is the amount spent on TV, radio, newspaper advertising.** * **Each column corresponds to the values of advertising budget for different media.** |
|  |
| ***Observation:***  **There are 200 Rows and 5 Columns in DataSet. TV, radio, newspaper, sales are floating data columns; unnamed: 0 is integer data type columns and the amazing part is there are no missing values in our Dataset.** |
| EXPLORATORY DATA ANALYSIS: Exploratory Data Analysis, or EDA, is an important step in any Data Analysis or Data Science project. EDA is the process of investigating the dataset to discover patterns, and anomalies (outliers), and form hypotheses based on our understanding of the dataset. |
| Let’s know more about our dataset: |
|  |
| ***Observation:***  **We can find some skewness in the datset, let’s visualize and confirm the same.** |
| DATA VISUALIZATION:  4 Reasons to Utilize Data Visualization Software | Inzata AnalyticsData visualization is the presentation of data in a pictorial or graphical format. It enables decision makers to see analytics presented visually, so they can grasp difficult concepts or identify new patterns.  Let’s plot and explore our dataset! |
|  |
| ***Observation:***  **Using distplot, found there are skewness in newspaper column.** |
| Now, let’s check the relationship between Features and the Target Variable |
|  |
| ***Observation:***   * **TV vs. Sales - Strong positive linear relationship i.e., as TV advertising is increased, sales are also increased, although, sometimes increasing the TV budget didn't increase sales.** * **The relationship between these two variables is approximately linear.** * **Radio vs. Sales - Weak positive relationship(comparatively) i.e., similar to TV advertising, as radio advertising is increased, sales are also increased. But Radio advertising is less effective than TV.** * **Newspaper vs. Sales - no relationship (we can discard such kind of variables which donot have a linear relationship) i.e., newspaper advertising has high effect on sales positively at initial. However, as advertising is increased, it has low effect on sales and almost nil at max.**   TV advertising has higher positive impact on sales and followed by radio. So, TV and radio will have higher correlation with sales¶ |
|  |
|  |
| **+1: positively correlated**  **-1: negatively correlated**  **0: no correlation**  ***Observation:***   * **We can see that there is no multi collinearity among independent/predictor variables** * **TV and Radio are highly correlated with sales data which coincides with the graphical data analysis done in above steps.** |
| ***Observation:***  **TV Advertising increases the sales high which is followed by Radio Advertising.** |
| EDA REMARKS:  **TV, Newspaper and radio advertising affect the target variable Sales. From the heatmap, we can observe that ‘Sales’ and ‘TV’ have a higher correlation as compared to others because it shows a linear pattern as well as giving high correlation.**  **So, as we saw that we have done a complete EDA process, getting data insights, feature engineering, and data visualization as well so after all these steps one can go for the prediction using machine learning model-making steps.**  **Usually, we go for feature engineering or feature selection steps after EDA. But we have fewer features and emphasis on actually using the model. So we are moving forward towards the next steps.** |
| PREPROCESSING PIPELINE: Concepts of Data Preprocessing – Brain Mentors |
| Data pre-processing is a data mining technique which is used to transform raw data into a useful format.  Let’s start building our model, firstly, we will seperate Features and Label Column. |
|  |
| As we have seen in the visualization, that “Newspaper” column contains skewness, let’s check the same.  Skewness - Overview, Types, How to Measure and InterpretSkewness:  It refers to a distortion or asymmetry that deviates from the symmetrical bell curve, or normal distribution, in a set of data. If the curve is shifted to the left or to the right, it is said to be skewed. |
| ***Observation:***  **After checking for the Skewness, we found that newspaper has more skewness as predicted by distplot analysis.**  **We will now perform “Power Transform” method to remove the skewness**. |
| ***Observation:***  **Rechecking the skewness after power transforms, we found that skewness of all columns are within range of -0.5 to +0.5** OUTLIER DETECTION: Introduction to Anomaly Detection in Python: Techniques and Implementation  | cnvrg.ioOutliers are observations in a dataset that don't fit in some way. An outlier is a data point that differs significantly from other observations. They may be due to variability in the measurement or may indicate experimental errors. |
|  |
| Next step is Data Standardization,  DATA STANDARDIZATION: It is the process of bringing data into a uniform format that allows analysts and others to research, analyze, and utilize the data. In statistics, standardization refers to the process of putting different variables on the same scale in order to compare scores between different types of variables. |
| BUILDING MACHINE LEARNING MODEL: Supervised learning — is a machine learning task that establishes the mathematical relationship between input X and output Y variables. Such X, Y pair constitutes the labeled data that are used for model building in an effort to learn how to predict the output from the input.  Let’s first understand the theory by Linear Model and then we will do Python Implementation: Linear Relationship A perfect linear relationship between an independent variable x and dependent variable y has the mathematical form:  y = β0 + β1 x.  iop β0 is called the -intercept and y is called the slope. Simple Linear Regression Simple linear regression is used to model the relationship between two continuous variables. Often, the objective is to predict the value of an output variable (or response) based on the value of an input (or predictor) variable. Recall Notation • m=number of training examples  • x =“input” variable/features  • y =“output” variable/“target” variable  • (x (i), y (i)) the i th training example  • x (1) = 230.1, y (1) = 22.1, x (2) = 44.5, y (2) = 10.4 |
| The best fit line for the data points is nothing but the line that best expresses the data point relationship. Residual Sum of Squares (RSS) is computed to find the best-fit line, such line will have the lowest value of RSS.  equation  In simple linear regression, if the coefficient of x is positive, it can be concluded that the relationship between independent variable and dependent variables is positive.  i.e., in Yi=ß0+ß1Xi +Ɛi  if ß1>0 the relationship is positive.  if ß1<0 the relationship is negative. Least Squares Vertical and perpendicular offsetsThe Least Squares approach is to find the y-intercept β0 and slope β1of the straight line that is closest to as many of the points as possible.  The least-squares explain that the curve that best fits is represented by the property that the sum of squares of all the deviations from given values must be minimum, i.e:  Least Square Method formula  Sum = Minimum Quantity  Suppose when we have to determine the equation of line of best fit for the given data, then we first use the following formula.  The equation of least square line is given by Y = a + bX  Normal equation for ‘a’:  ∑Y = na + b∑X  Normal equation for ‘b’:  ∑XY = a∑X + b∑X2  Solving these two normal equations we can get the required trend line equation.  Thus, we can get the line of best fit with formula y = ax + b  The estimated linear regression of sales on newspaper is:  yi = 12.35 + 0.05xi,  where yi is sales in ith the market and xi is the dollar amount spent on newspaper  advertising in the ith market.   * The slope β̂1 is the amount of change in for a unit change in x. * Sales increase by 0.05 for each dollar spent on advertising. * The intercept β̂ 0 is the average of y when xi = 0. * The average sales are 12.35 when the amount spent on advertising is zero.  Prediction using a Linear Regression Model Prediction is determined by the value of the variable. Accuracy and fitness is measured by loss, R square, adjusted R square etc. After a linear regression model is estimated from data it can be used to calculate  predicted values using the regression equation  ŷi = β0 + β1xi.  ŷi is the predicted value of the ith response yi.  The ith residual is:  ei = yi − ŷi    Gradient Descent  It is an optimization algorithm used for minimizing the cost function in various machine learning algorithms. Linear Regression With Multiple Variables:Advertising Dataset:  |  |  |  |  | | --- | --- | --- | --- | | TV  *x*1 | Radio  *x*2 | Newspaper  *x*3 | Sales  *y* | | 230.1 | 37.8 | 69.2 | 22.1 | | 44.5 | 39.3 | 45.1 | 10.4 | | 17.2 | 45.9 | 69.3 | 9.3 | | . | . | . | . |        Gradient Descent with Multiple Variables:   So, it’s time for Python implementation, Let’s begin with spliting our Dataset into Train and Test Data; Train-Test-Split: The train-test split is a technique for evaluating the performance of a machine learning algorithm. It can be used for classification or regression problems and can be used for any supervised learning algorithm. The procedure involves taking a dataset and dividing it into two subsets.    Now let’s prepare the data to fit into a machine learning model and then we will use different regression algorithms to train a sales prediction model using Python. |
|  |
|  |
|  |
| After training different models, we found that Lasso Regression Model has the least difference in r2\_score so it is the best model. We will now improve the accuracy of the model with Hyper parameter tuning.  Hyperparameter tuning: It is choosing a set of optimal hyperparameters for a learning algorithm. A hyperparameter is a model argument whose value is set before the learning process begins. The key to machine learning algorithms is hyperparameter tuning. |
|  |
| ***Observation:***After Hyper parameter Tuning, Lasso Model r2\_score is increased from 83% to 98.53% Now, the Final Step is to save the model for Future Predictions: |
| CONCLUDING REMARKS:  * Increase Sales GIFs - Get the best GIF on GIPHYTV, Newspaper and radio advertising affect the target variable Sales. The order of correlation with sales given by TV Adv> Radio Adv> Newspaper Adv. * The linear model created with TV, Newspaper and radio as input variable and Sales as Target variable gave the higher r2score. * The linear model created with TV as input variable and Sales as Target variable gave the second highest r2score. |
| * The linear models created with Radio and Newspaper as input variables and Sales as Target variable has very less r2 score. * Therefore, in order to create a suitable model which approximates the scenario correctly, we need to use the combination of TV, Radio and Newspaper as target variable. * It is clear from the model and correlation matrix that Amount spend on TV ads directly impact the Sales of a particular product. * We can conclude from the R2 value that only input variable TV advertisements provides a satisfactory result.  About the Author: Pursuing Post Graduate Diploma in Data Science, Machine Learning and Neural Network. I am an aspiring Data Scientist whose purpose is to learn in detail all the concepts needed for Data Science. I am passionate about Data Science and have skills that help me derive valuable insights from data, such as Data Manipulation, Data Visualization, Data Analysis, EDA, and Machine Learning. |
|  |
|  |