**Sales revenue**

**Problem statement**

This case study would help in understanding how advertisement spends done by a company for their

product has helped in driving their Sales Revenue and also help us in planning the future

advertisement spends. We have two variables in this dataset,

AdSpends = The advertising expenditure (in thousands of dollars)

SalesRevenue = The sales revenue (in thousands of dollars)

**Aim**

To find the predecion of sales revenue with Adspend

**Solution**

Here the problem is to find the Patten sales revenue with the help of adsend

There is only one attribute which are Ad spend

Independent value - Sales revenue

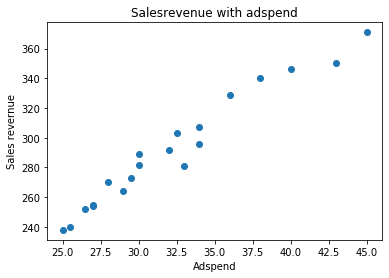
Dependent value - Ad spend

1. Find the 5 point summary of the dataset.

* The data set is really small so we can’t use any high level model to predict
* Don’t have any missing values to impute
* Ad spend is little skewed
* Both the variable are has strong correlation it’s good to for model prediction
* It seems the is a positive trend on adspend and sales when analysis with help of RETURN ON AD SPEND from certain point to starts to fall
* When the business aim is to maximise revenue for advertising expenditure, the ROAS metric is very useful. It is assumed business performance is driven by absolute return on advertising spend and as such the ROAS metric is targeted.

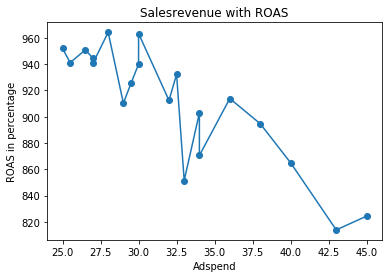
2. Plot the dataset to see how the data is distributed.

**Sales revenue with Ad spend**

****

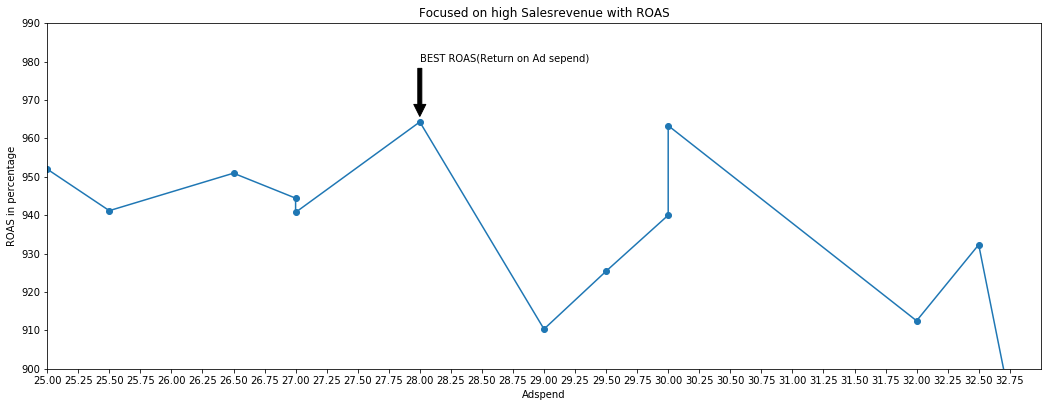
It seems the is a positive trend but we have analysis with help of RETURN ON AD SPEND.

* When the business aim is to maximise revenue for advertising expenditure, the ROAS metric is very useful. It is assumed business performance is driven by absolute return on advertising spend and as such the ROAS metric is targeted.

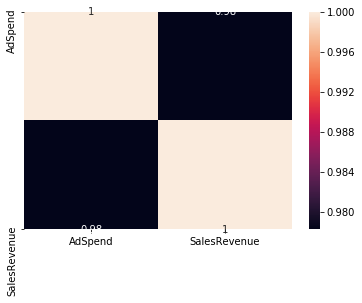


It clearly show that there after some point there is drop in ROAS(return on Adspend) the higher point with ranges form to to 27.5 to 30

**Focused on high Salesrevenue with ROAS is between $28.00 to $28.25**



3. Determine how the variables are related to each other. If there is any relation explain it.



There is a strong correlation between the variables is good to gave

|  |  |  |
| --- | --- | --- |
|  | **ad spend** | **sales revenue** |
| **ad spend** | **1.00** | **0.978282** |
| **sales revenue** | **0.978282** | **1.00** |

* Ttest result is also good so these are statistically good samples for prediction

4. Build the linear regression model and interpret all the statistics obtained from the model.

R square value for liner model = 0.96

Coefficient = 224.5671

RMSE of Train = 7.140603

RMSE of TEST = 7.070045

Also justify your conclusions about the model.

The model has done a good job with less RMSE and spening money more tha 28.00 will decrease the Return form Ad spend to the best fit to spend is $28.00(in thousands of dollars)

5. Perform all the model diagnostics to check if all the assumptions are satisfied.

Data is more small cant use any complex algorithms Best fit is liner model

There are four major assumptions associated with a linear regression model:

1. **Linearity**: First, linear regression needs the relationship between the independent and dependent variables to be linear.  The linearity assumption can best be tested with scatter plots it show that these data is liner.
2. **Heteroscedasticity**: There is only one dependent variable and as per the data there is less Homoscedasticity.
3. **Multicolniarity**: Observations are independent of each other. Here we have only one variable so there is no chance of multicolniarity.
4. **Auto correlation**: There is no auto correlation.
5. **Endogenity** : yes there is only one dependent variable if there is some more independent variable the prediction will be more strong
6. Use the model built to predict for an AdSpend = 50

If the company spend $50 we will get the revenue of $392.97(in thousand dollars)