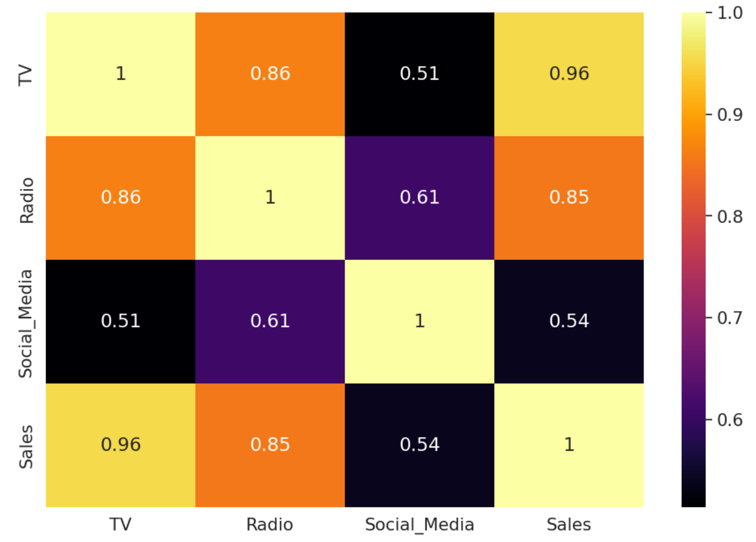
Market Mix Modeling

Title: Performing Market Mix Modeling on an advertising data.
We have collected the data from an advertising firm which tells us how many dollars were spend on different platforms through advertising.
We are going to analyze the data graphically and numerically to understand that which platform have given us more returns, which platform is stable and monotonous. Finally, we will be performing an optimisation process to check the spending ratios, when advertising dollar has been contrained.
For this project, we're going to use a dataset that consists of marketing spend on TV, radio, and Social media, as well as the corresponding dollar sales by period.

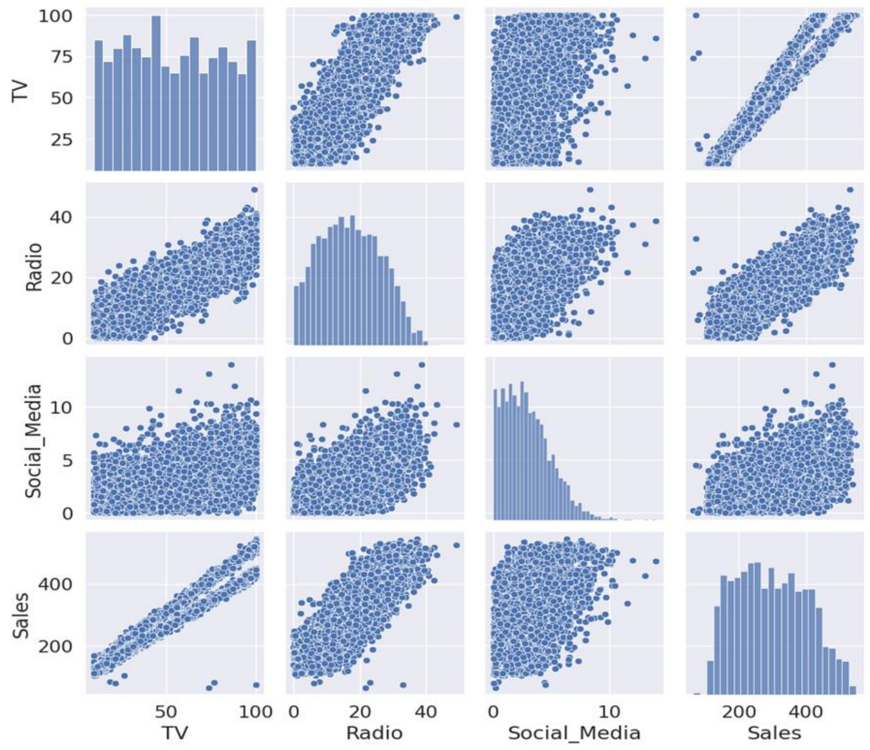
Date	TV	Radio	Social_Media	Influencer	Sales
2001-01-01	54	14.119060	4.215642	Nano	324.905658
2001-01-02	65	26.217056	5.481302	Macro	370.511432
2001-01-03	67	16.705375	2.657608	Macro	389.325067
2001-01-04	73	31.327127	10.260554	Nano	412.936802
2001-01-05	28	17.343639	2.071636	Micro	214.546227

Checking Correlation



From the above plot we can say that initially there is a higher positive correlation between TV and Sales (0.96) than between Radio and Sales (0.85) or Social Media and Sales (0.54)

Pair Plot



The above plot shows there is a flat correlation between Social Media and Sales, whereas correlation between Radio and Sales is a bit better. And correlation between TV and sales is the best but there are some outliers we can notice.

Linear Regression

```
                            OLS Regression Results
==============================================================================
Dep. Variable:                  Sales   R-squared:                       0.917
Model:                            OLS   Adj. R-squared:                  0.917
Method:                 Least Squares   F-statistic:                 1.685e+04
Date:                Tue, 28 Nov 2023   Prob (F-statistic):               0.00
Time:                        03:08:53   Log-Likelihood:                -21977.
No. Observations:                4552   AIC:                         4.396e+04
Df Residuals:                    4548   BIC:                         4.399e+04
Df Model:                           3
Covariance Type:            nonrobust
================================================================================
                   coef    std err          t      P>|t|      [0.025      0.975]
--------------------------------------------------------------------------------
Intercept       85.1463      1.052     80.904      0.000      83.083      87.210
TV               3.4555      0.034    101.424      0.000       3.389       3.522
Radio            0.9832      0.105      9.385      0.000       0.778       1.189
Social_Media     2.4691      0.281      8.774      0.000       1.917       3.021
==============================================================================
Omnibus:                     1076.006   Durbin-Watson:                   0.572
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            18834.161
Skew:                          -0.663   Prob(JB):                         0.00
Kurtosis:                      12.876   Cond. No.                         148.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

Summary provides us with an abundance of insights on our model. Two main things that are most useful for us in this.

1. The Adj. R-squared is 0.917, which means that almost 92% of all variations in our data can be explained by our model, which is pretty good.

2. The p-values for TV, radio and Social_Media are 0.000, which indicates that they have a significant impact on sales.

From the values above we are getting this formula:
Sales = 85.1463 + 3.4555 * TV + 0.98 * Radio + 2.469 * Social Media

Interaction effect between each Variables

```
TV                     0.000000e+00
Radio                  4.894864e-36
Social_Media           5.894242e-39
TV:Radio               4.270327e-99
TV:Social_Media        5.265578e-28
Radio:Social_Media     1.603664e-11
dtype: float64
```

As we can see there is a presence of interaction terms. Also, interaction with TV and Radio, TV and Social Media, Radio and Social Media proves to be very important for the prediction of Sales as we are getting p-values close to 0 for each interaction.

Breaking Down the Sales

To illustrate the computation of the contributions, let us consider a single week:

	TV	Radio	Social_Media	Influencer	Sales
Date					
2001-07-16	84	27.817582	6.171058	Micro	450.741646

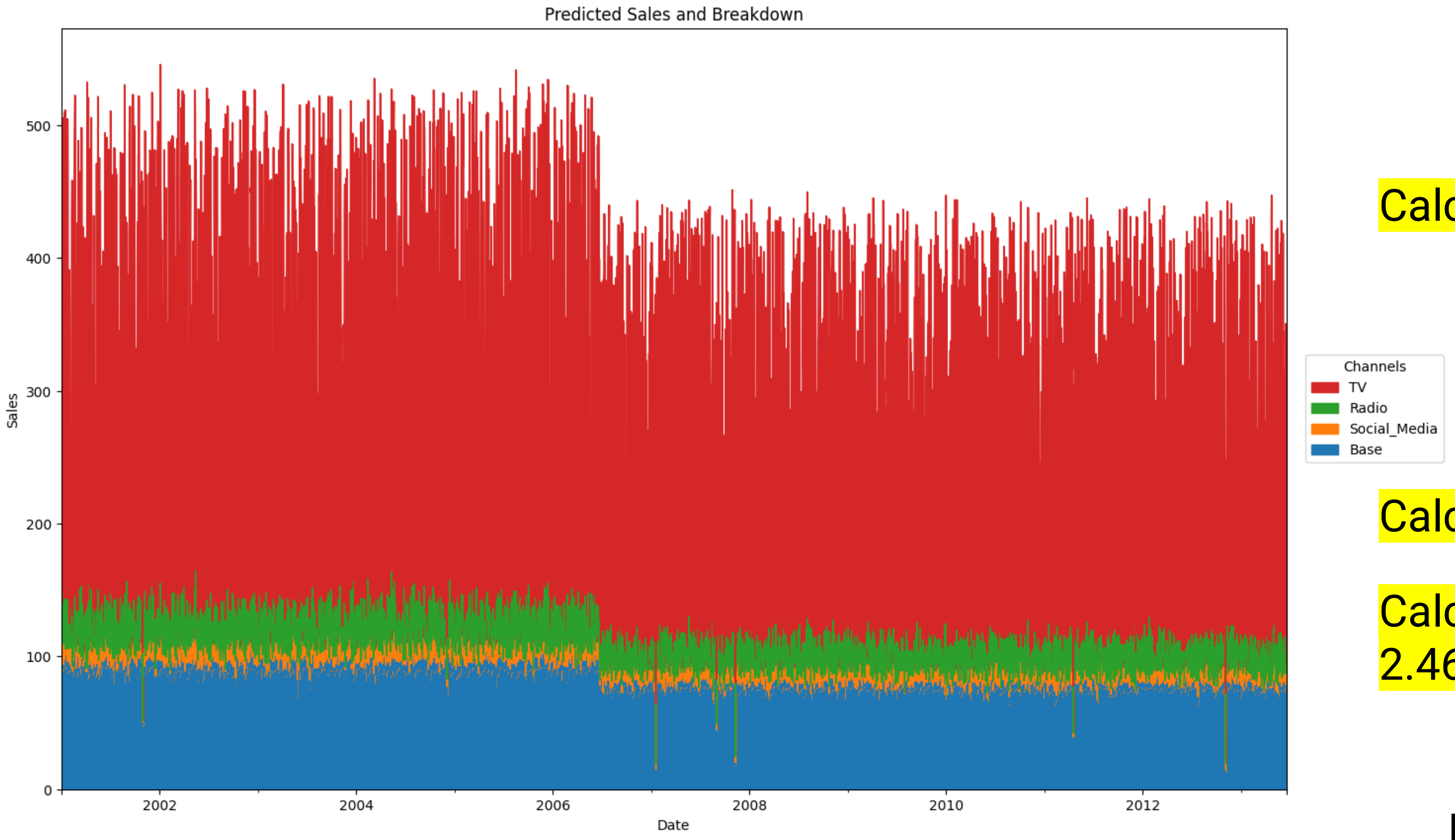
Prediction

```
print(lr.predict(([[84.0, 27.817582, 6.171058]])))
```

```
[417.99253206]
```

This is not exactly the true answer of 450.741646 from the table above, but let us stick with it for now. We can now see that the (unadjusted) contribution of TV is

Contribution plot for all observations



We can see (or compute) that the baseline is around 80 sales each day, Social Media contributes around 80-100 on average whenever it is active, Radio contributes around 100-150 on average whenever it is active , and TV contributes around 150- 450 when it is active

Return on Investment (ROI)

Calculating the TV ROI : 3.456317384842416

An ROI less than 1 means that the channel performed poorly. For the TV ROI we can say : For each 1 $ we spent on TV, we got 3 dollars 46 cents back.

Calculating the Radio ROI :0.9822807902564936

Calculating the Social Media ROI : 2.465709000491316

From the above results we can see, TV and Social Media has an ROI of 3.46 and 2.47 which is much better, it seems that these channels worked quite well in the time period that we considered.

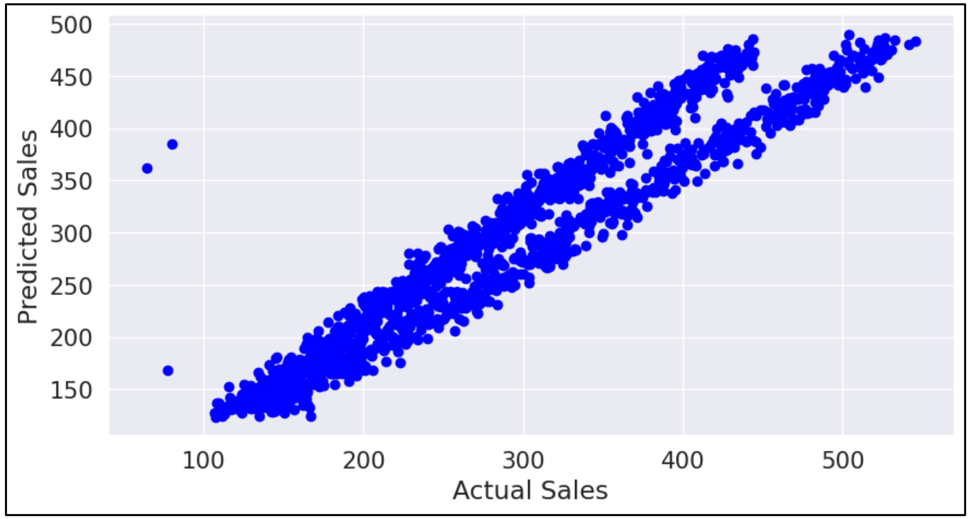
Created dummy variable and ran a linear regression and got these coefficients:

	Coefficient
TV	3.454105
Radio	0.975709
Social_Media	2.612431
Macro	2.629586
Micro	0.738496
Nano	1.543664

Interpreting the Coefficients

If we take a closer look above we understand that the spends on TV have a positive causal relationship with sales, as an increase of 1 Mn dollars in TV spends has an effect of 3.45 Mn increase in Sales. It is however interesting to note that Radio and Social media also have a positive impact on Sales as they increase the sales by 0.97 Mn and 2.61 Mn dollars with every additional 1 Mn dollars spent.
Now since Macro, Micro and Nano are binary variables, what their coefficients mean is that whenever the variable is flagged 1, i.e. suppose a Macro influencer was used, sales increased by around $2.63 Mn.

Actual Sales V/S Predicted Sales



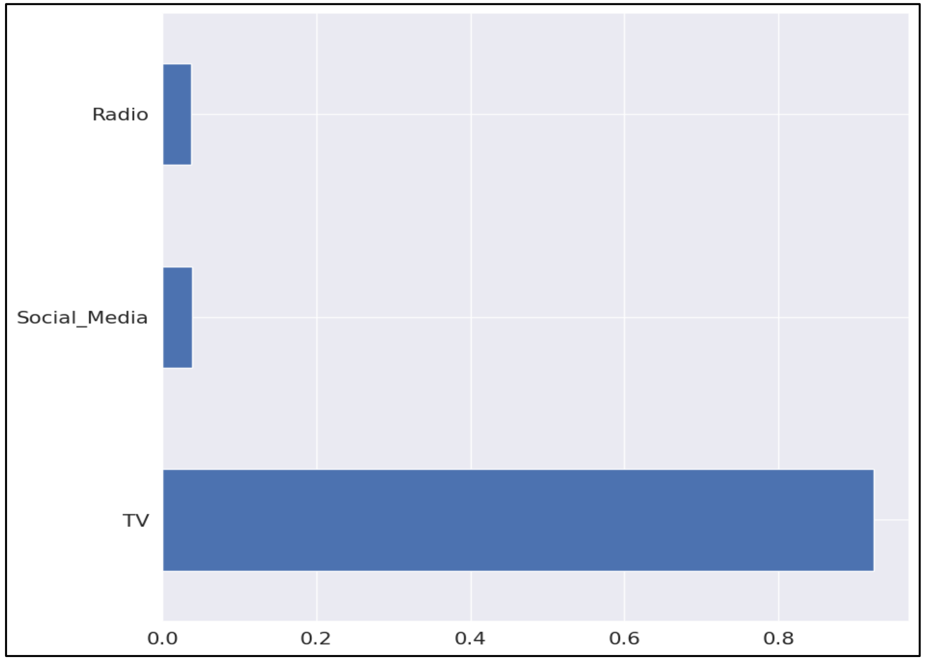
The above chart shows us that the predicted values are less than the actual values, means it is a little bit underfitting

Evaluating the Model

MAE: 25.915060314272694
MSE: 972.1516530307853
RMSE: 31.17934657799591
R-Square: 0.9126916704947571

- R-Square of 0.91 implies that the independent variables i.e. the marketing spends across the channels are able to successfully explain 91% of the variability seen in Sales across the dataset.
- Mean absolute error (MAE) is calculated as the mean of the absolute difference in the predicted sales of the test dataset and the actual values of Sales given. Here, there is a fluctuation of about 25.915 Mn dollars between the predicted and actual values of Sales.
- Root Mean Square error (RMSE) is another transformed variation of measuring the fluctuations between actual and predicted sales. It is found to be $31.17 Mn, which is also close to the MAE reported.

Feature importance with Random Forest




```
TV              0.923749
Radio           0.038017
Social_Media    0.038233
```

There seems to be a pattern, where TV is the most important, followed by Social Media, leaving radio last.

Printing each and every regressors with run time, rmse and rmse_cv

rmse = root mean square error
rmse_cv = root mean square error cross validation

	model	run_time	rmse	rmse_cv
0	XGBRegressor	0.005882	31	30
15	BayesianRidge	0.0016767	31	30
14	ARDRegression	0.0042899	31	30
12	HuberRegressor	0.0034314	31	30
10	Lars	0.0016646	31	30
8	LinearRegression	0.0019095	31	30
9	Ridge	0.0020752	31	30
17	OrthogonalMatchingPursuit	0.0015006	32	31
1	RandomForestRegressor	0.2036737	32	31
6	LinearSVR	0.0007787	35	33
18	pls	0.0013018	35	33
13	PassiveAggressiveRegressor	0.000945	38	33
16	ElasticNet	0.0012618	41	40
2	DecisionTreeRegressor	0.0039741	43	41
4	SVR	0.1013323	43	42
11	TheilSenRegressor	0.1824372	45	42
5	NuSVR	0.0544184	44	44
7	KernelRidge	0.1046952	298	298
3	GaussianProcessRegressor	0.2909849	4877	10011

The first outcome i.e. Ridge Regressor is the best model for out dataset
Now predicting y value with help of x_test.

Actual and Predicted values

This model works, but it's only crude given the complexity of many MMMs. To further improve this, there are a whole load of other things we could do. Firstly, we could add some lagged data, by using shift to get the channel spend in the previous days to see if that's correlated to sales in future periods. We could also do some hyperparameter tuning, or add in some additional regressors if we have them.

Carry-Over Effect :

This one is slightly more tricky. Let us use an example to understand what we want to achieve. We are given a series of spendings over time such as (16, 0, 0, 0, 0, 4, 8, 0, 0, 0), meaning that we spent 16 in the first week, then we spent nothing from week 2 to 5, then we spent 4 in week 6, etc.

We now want that spendings in one week to get partially carried over to the next weeks in an exponential fashion. This means: In week 1 there is a spend of 16. Then we carry over 50%, meaning

$0.5 * 16 = 8$ to week 2,

$0.5^2 * 16 = 4$ to week 3,

$0.5^3 * 16 = 2$ to week 4,

This introduces two hyperparameters:

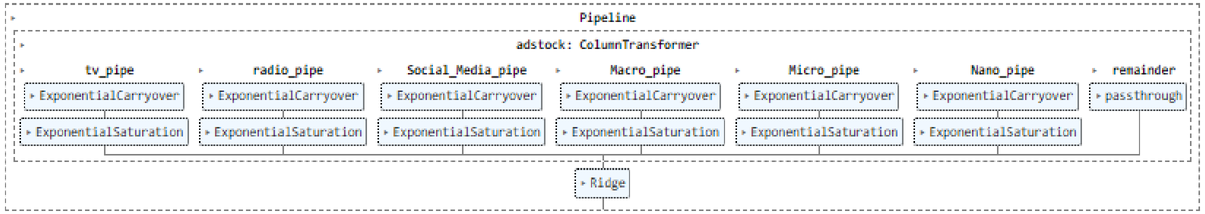
The strength (how much gets carried over?) and the length (how long does it get carried over?) of the carry-over effect. If we use a strength of 50% and a length of 2, the spending sequence from above becomes (16, 8, 4, 0, 0, 4, 10, 5, 2, 0).

Advertising Adstock :

This feature engineering that we will do is a crucial component called the advertising adstock, a term coined by Simon Broadbent. It is a fancy word that encapsulates two simple concepts:

1. We assume that the more money you spend on advertising, the higher your sales get. However, the increase gets weaker the more we spend. For example, increasing the TV spends from 0 € to 100,000 € increases our sales a lot, but increasing it from 100,000,000 € to 100,100,000 € does not do that much anymore. This is called a saturation effect or the effect of diminishing returns.
2. If we spend money on advertising week T, often people will not immediately buy your product, but a few (let us say x) weeks later. This is because the product might be expensive, and people want to think about it carefully or compare it to similar products from other companies. Whatever the reason might be, the sale in week T + x is partially caused by the advertising you played in week T, so it should also get some credits. This is called the carry-over or lagged effect

The Final Model

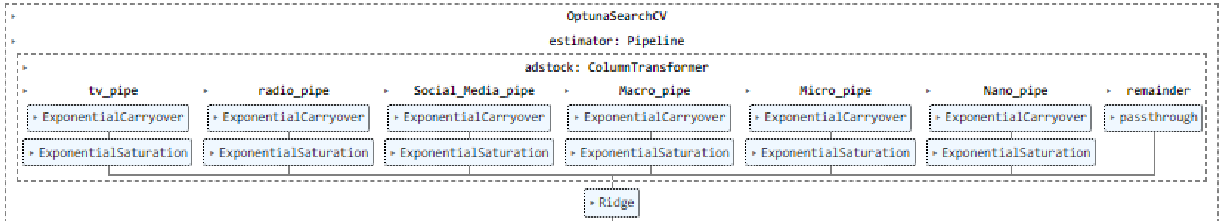


Cross Validation Score : 0.017147981656106516

Hyperparameter Tuning :
We will use Optuna, an advanced library for optimization task. Among many other things, it offers a scikit-learn-compatible OptunaSearchCV class that we can see as a drop-in replacement of sckit-learn's GridSearchCV and RandomizedSearchCV .
In a nutsell, OptumaSearch is a much smarter
version of RandomizedSearchCV. While RandomizedSearchCV walks around randomly only, OptunaSearchCV walks around randomly at first, but then checks hyperparameter combinations that look most promising.

Cross Validation Score : 0.86

Again fit the model with Hyperparameter Tuning



- Printing best Parameter value for each X variable in a DataFrame

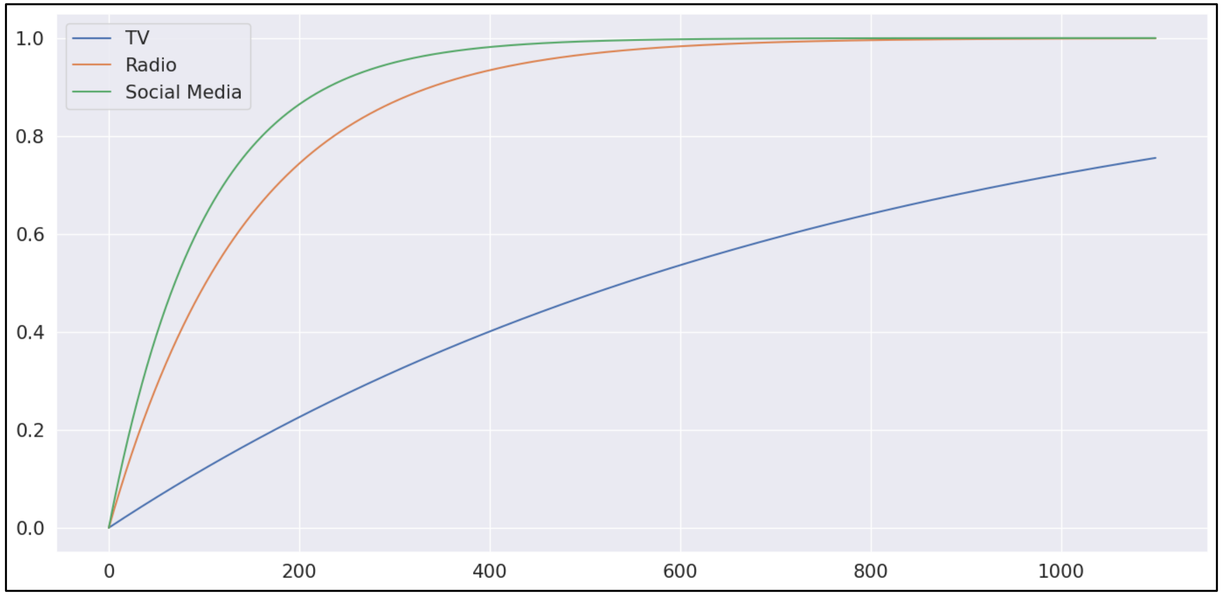
	value
adstock__tv_pipe__carryover__strength	0.730899
adstock__tv_pipe__carryover__length	0.000000
adstock__tv_pipe__saturation__a	0.001281
adstock__radio_pipe__carryover__strength	0.348483
adstock__radio_pipe__carryover__length	4.000000
adstock__radio_pipe__saturation__a	0.006809
adstock__Social_Media_pipe__carryover__strength	0.609778
adstock__Social_Media_pipe__carryover__length	6.000000
adstock__Social_Media_pipe__saturation__a	0.010000
adstock__Macro_pipe__carryover__strength	0.819740
adstock__Macro_pipe__carryover__length	4.000000
adstock__Macro_pipe__saturation__a	0.005173
adstock__Micro_pipe__carryover__strength	0.081295
adstock__Micro_pipe__carryover__length	5.000000
adstock__Micro_pipe__saturation__a	0.006169
adstock__Nano_pipe__carryover__strength	0.547693
adstock__Nano_pipe__carryover__length	2.000000
adstock__Nano_pipe__saturation__a	0.001509

We can see from the above results, TV has the most carryover strength while Social Media & Radio have most carryover length. It means TV ads have high impact on Sales after initial spending, but they don't last long, quickly wear off in the same week. While Social Media and Radio have relatively less impact on Sales but they still have an effect on sales respectably 4 & 6 weeks later. Interestingly Macro influencers have both high carryover length and strength.

Saturation Effect

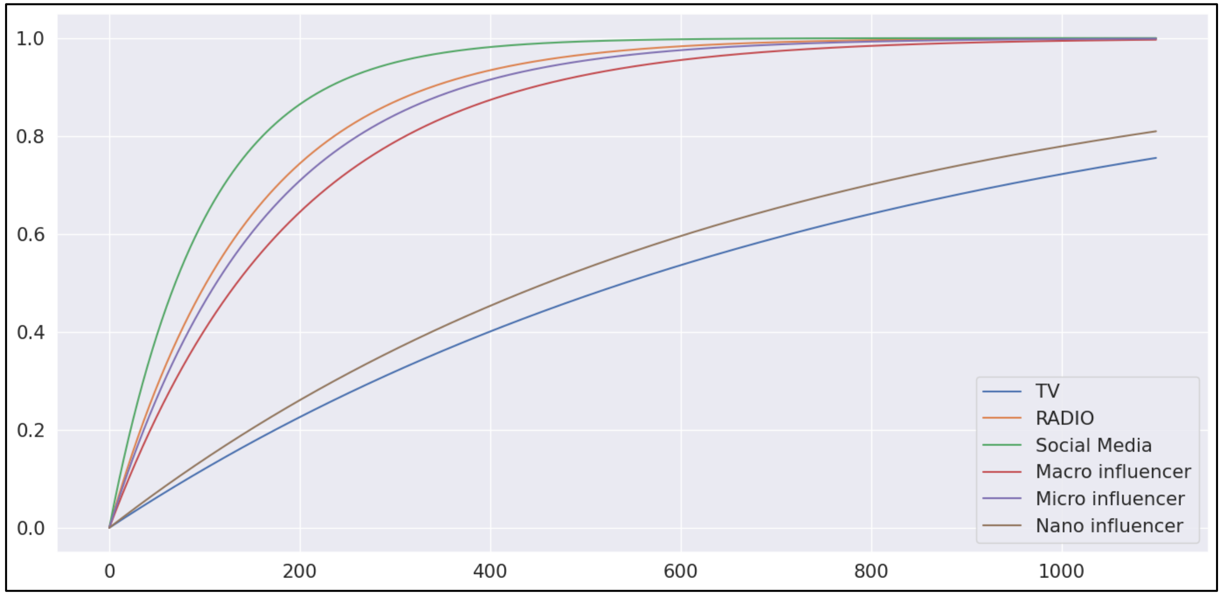
We want to create a transformation (=mathematical function) with the following properties:
1. If the spendings are 0, the saturated spendings are also 0.
2. The transformation is monotonically increasing, i.e. the higher the input spendings, the higher the saturated output spendings.
3. The saturated values do not grow to infinity. Instead, they are upper bounded by some number, let us say 1.

Without Selecting the Dummy Variables



We are getting saturation graph of tv radio and social media in same plot it is making sense to compare their saturation graphically.
From here it seems like spending more than 300 is useless for Radio and spending more than 230 is useless for Social Media

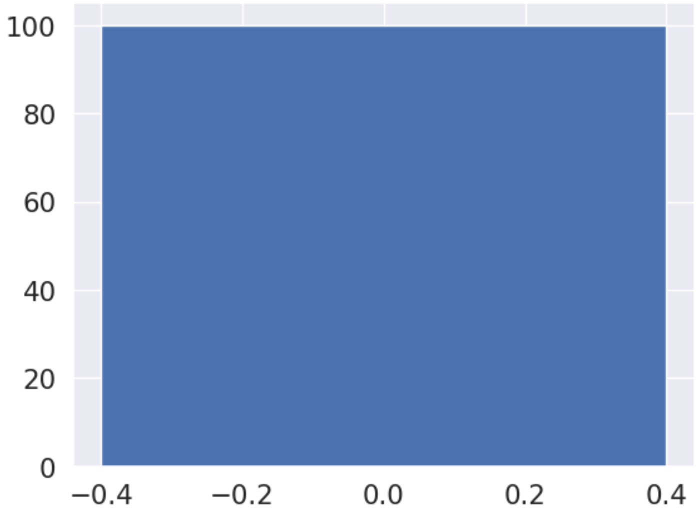
With using the Dummy Variables



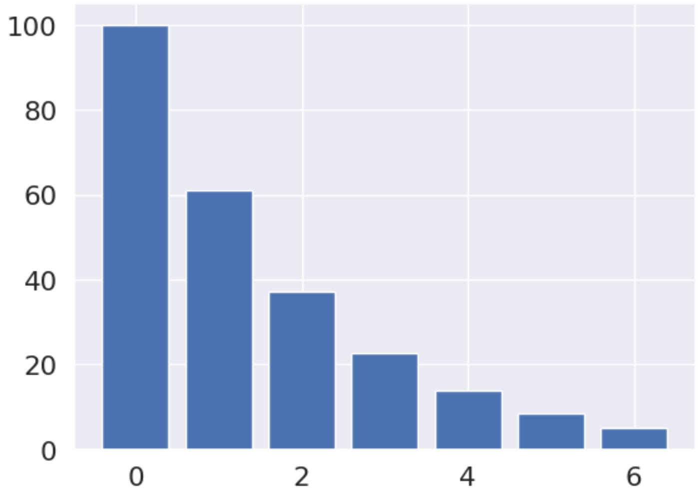
In the above chart, we are getting saturation graph of tv, radio, social media and other dummy influencers.

Carryover Effects

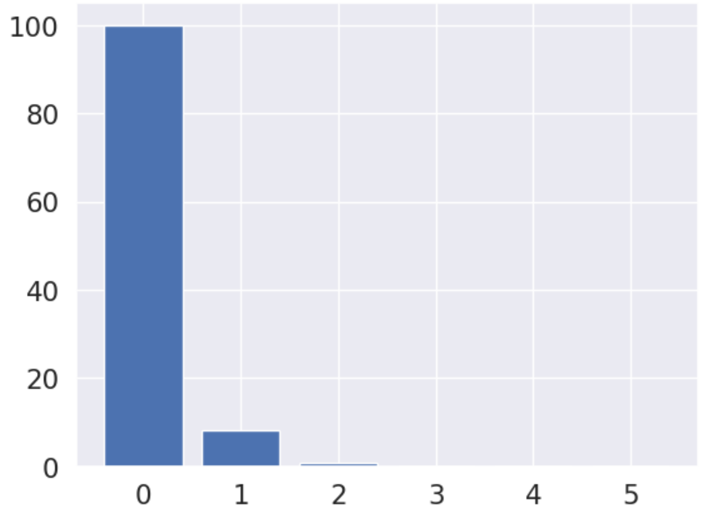
For TV



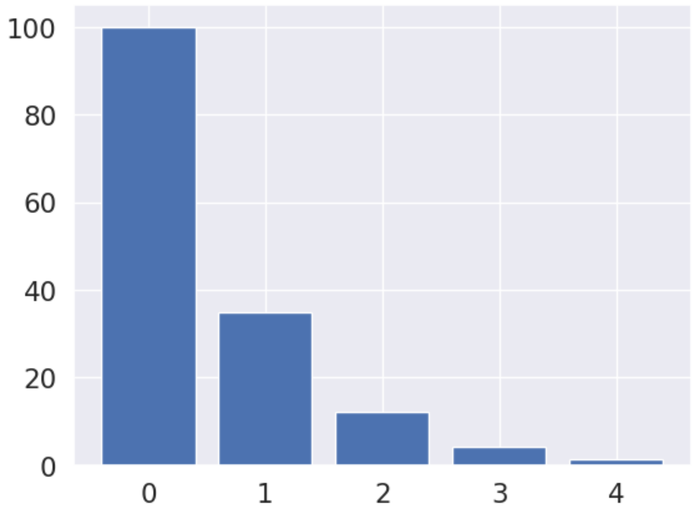
For Social Media



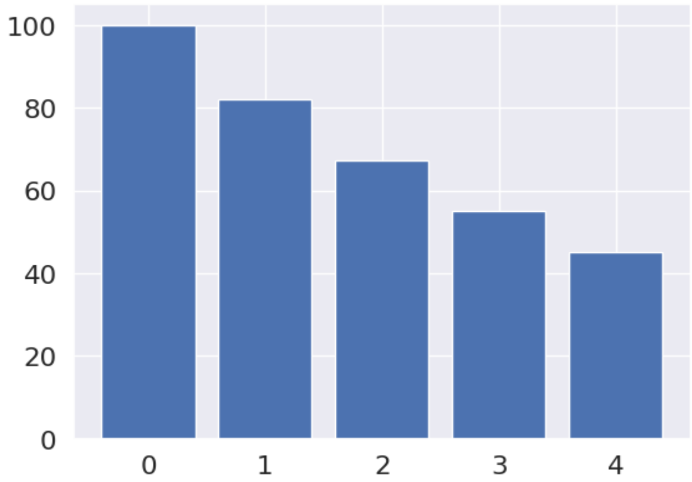
For Micro Influencers



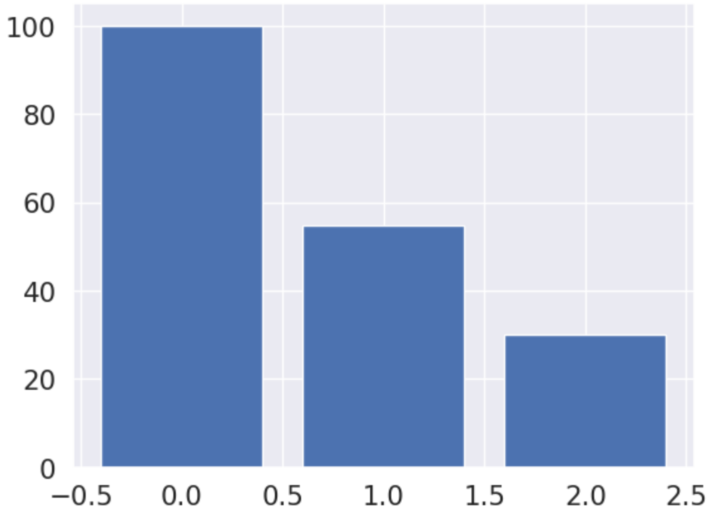
For Radio



For Macro Influencers

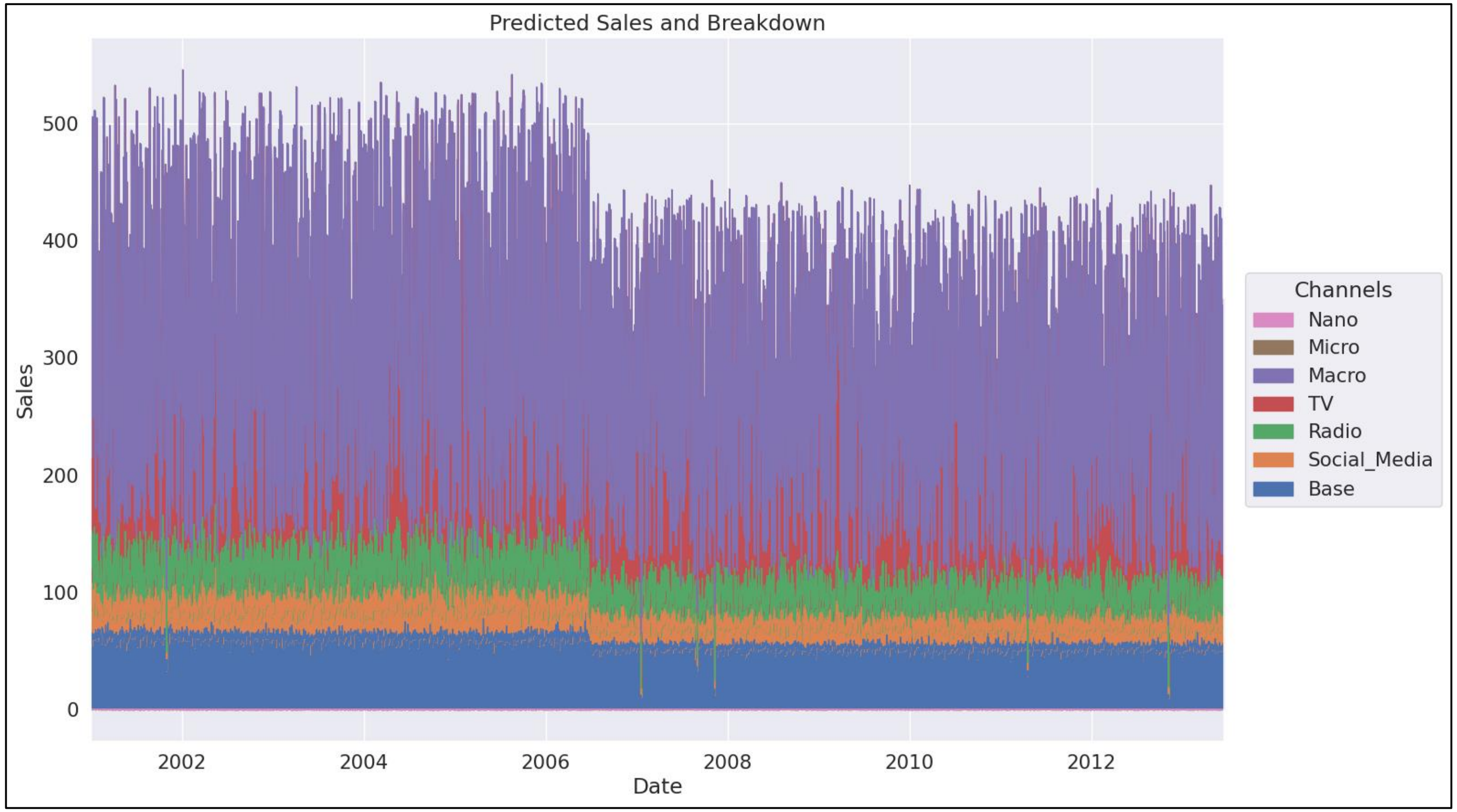


For Nano Influencer



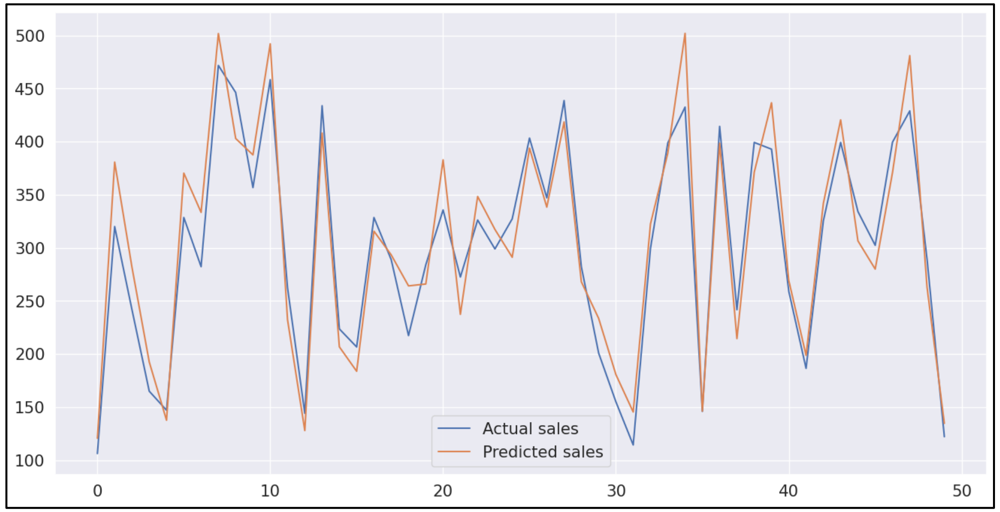
It seems that radio and Social Media advertisings still have an effect on sales respectably 4 & 6 weeks later after the initial spending. This is longer than the effect of TV spendings that quickly wear off in the same week.

Channel Contributions



It seems Macro influencers have an higher impact on sales, even greater than TV which additionally still has an effect on sales 4 weeks later after initial spending.

- Comparing Actual and Predicted values with tuned Model



- Checking For Multicollinearity using VIF values of the Explanatory Variables

```
TV              3.938138
Radio           4.592562
Social_Media    1.588171
Macro           1.485022
Micro           1.493019
Nano            1.489887
Constant        8.405170
dtype: float64
```

From the above results we can clearly see there is no Multicollinearity issue between the independent variables (all values are less than 5)

Using docplex to implement Solver in Python

Our Simple Linear Regression equation for the original data was:
Sales = 85.1463 + 3.4555 * TV + 0.98 * Radio + 2.469 * Social Media

Constraint on total ad Spent : $100000

My Linear Regression Coeffiecients

TV Coef = 3.4555
Radio Coef = 0.98
Social Media Coef = 2.469
intercept = 85.1463

Optimized Budget

Maximizing the Sales/Revenue through the intercept and coefficients of linear regression

(TV*TV_coef + Radio*Radio_coef + SM*SM_coef + intercept)

Using Docplex Model Solver we got Result :

TV = 99780
Radio = 120
Social_Media = 100

So, based on the results, if we have $100000 budget and we want to maximize our return of investment or objective, we should allocate 99780 dollars on Tv advertisements, 120 dollars on Radio advertisement and 100 dollars on Social Media advertisement.