CASE STUDY: MEDIA BUDGET ALLOCATION FOR BEAUTY BRANDS

Author: Shafel Mc Dowall

Objective:

- **Determine the best allocation of the total media budget among the 5 brands to maximize overall return on investment, based on mathematical criteria.**

Problem formulation:

The total media budget is set to €22,500,000 and must be allocated across the 5 brands for the 12 months of 2023.

The budget must be optimized with respect to the ROI which is directly calculated using the revenue column. More specifically: ROI=(Sales Growth - Marketing Cost) / Marketing Cost, where Sales Growth is equal to the difference between current and previous revenue.

Given the limited data and information about the provided data, a mixed model approach was used to provide the best possible outcome.

Assumptions:

- Market share can be calculated as total revenue of 5 projects. Data was cross referenced by calculating the difference from market share and total share held by competitors
- Sales Growth = Current revenue - previous revenue
- ROI = (Sales Growth - Marketing Cost) / Marketing Cost
- First ROI (2020-02-01) is equal to media budget to avoid creating outliers
- Return on ad spend is Revenue/Media Budget i.e. revenue per each euro spent on advertising
- Return on ad spend = Media Budget/Revenue, i.e the ratio of revenue to ad-spend

Caveats:

- The variable *'Monthly contribution to Market Share'* could not be ascertained from the data provided. Although the total for all 5 products per month equated to 100% it did not match the associated revenue, therefore, this value was excluded from the models.
- Other variables such as '*Budget to Revenue Ratio*', '*ROI*' and, '*Return on ad Spend*' were also excluded from the statistical models. As these variables are all directly calculated using revenue, including them in the model would result in a collinearity trap and contribute to overfitting or spurious results.
- Three models were used in the final solution to accommodate the lack of data and information.

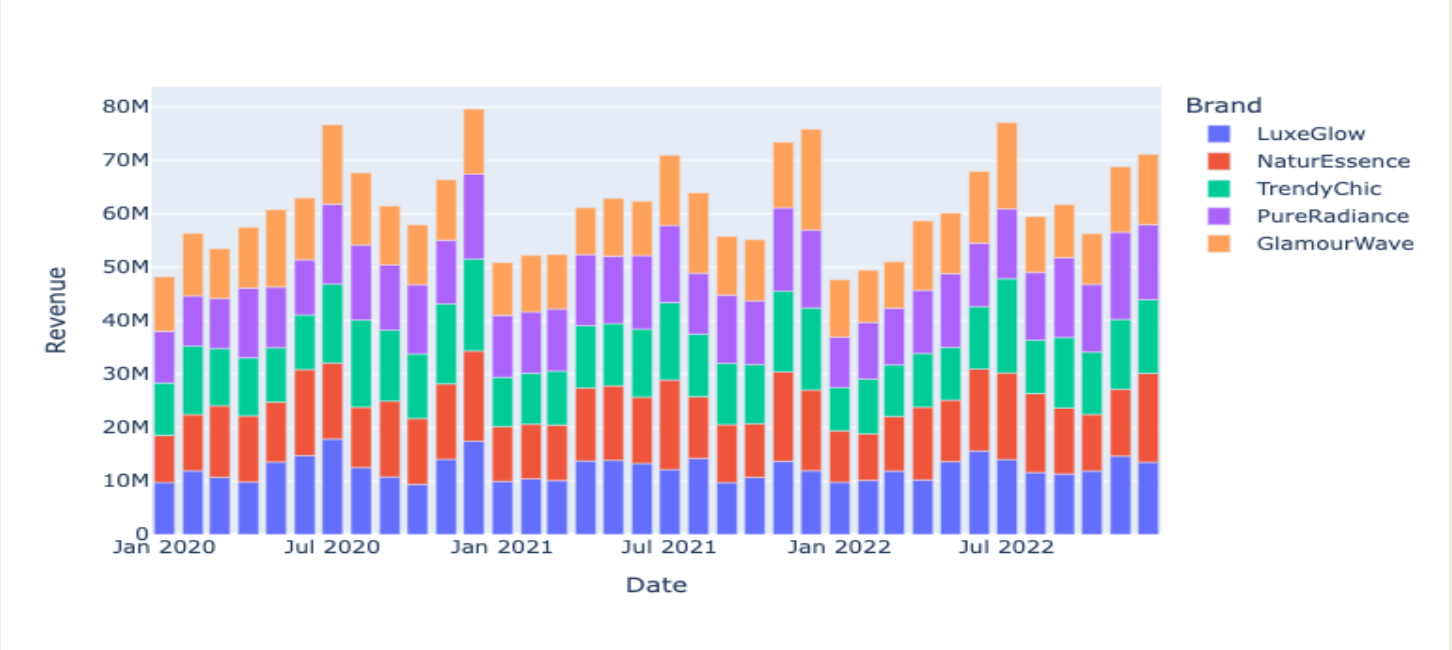
EXPLORATORY DATA ANALYSIS

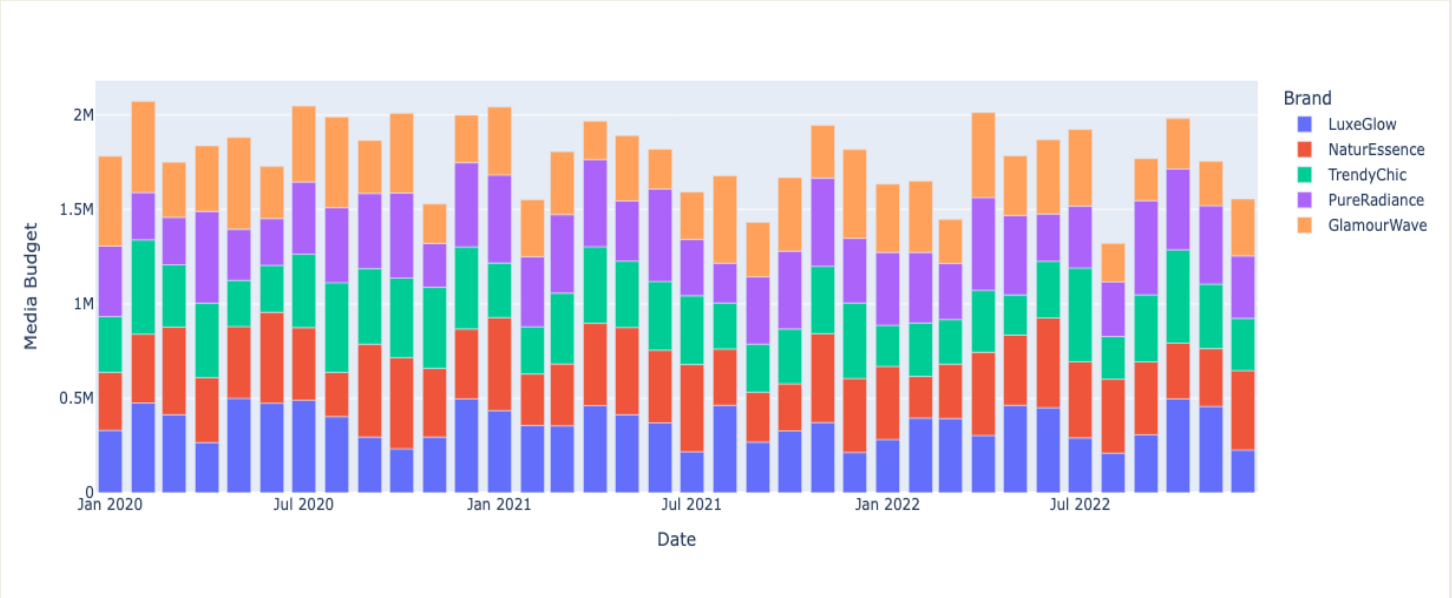
Descriptive Statistics

	Media Budget	Revenue
count	180	180,00
mean	357911,1111	12309710,00
std	87327,11821	2220322,00
min	203000	8091780,00
25%	288750	10453360,00
50%	364000	11897570,00
75%	428250	13791790,00
max	500000	18866700,00

- The Revenue has a standard deviation of 2million and an average of 12million, this is a large difference and highlights potential outliers.
- Conversely, the media budget does not have as great of a difference between the mean and std.

Note: Only samples of the EDA will be shown in the ppt to avoid overwhelming the user with slides.

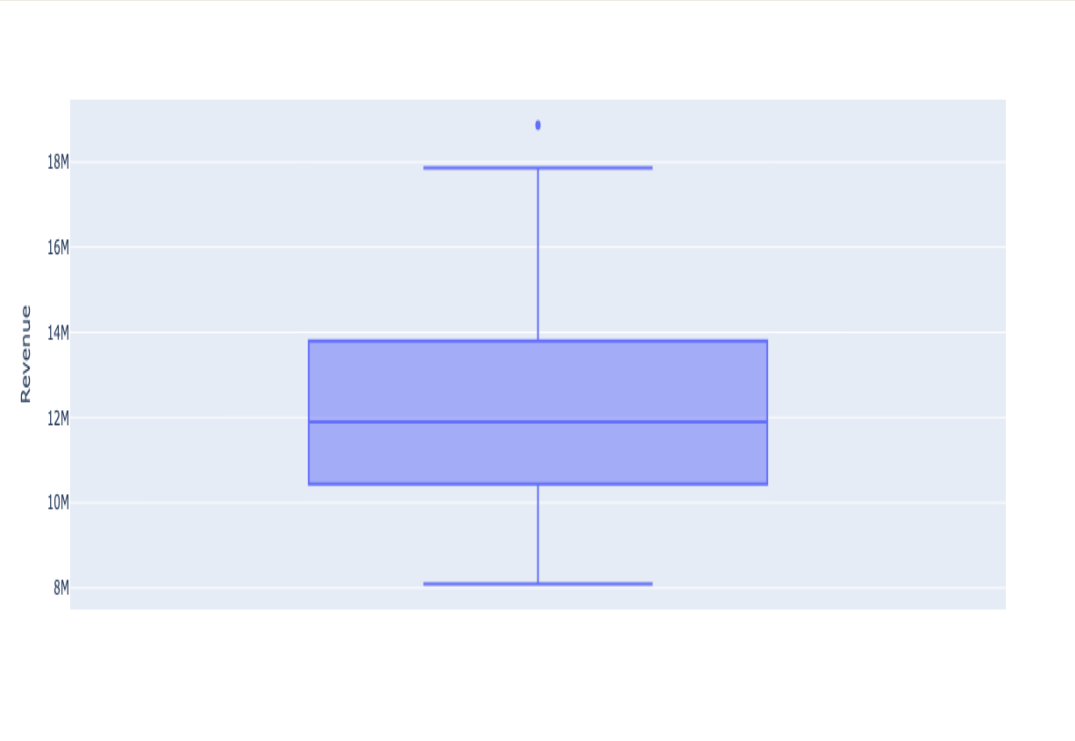


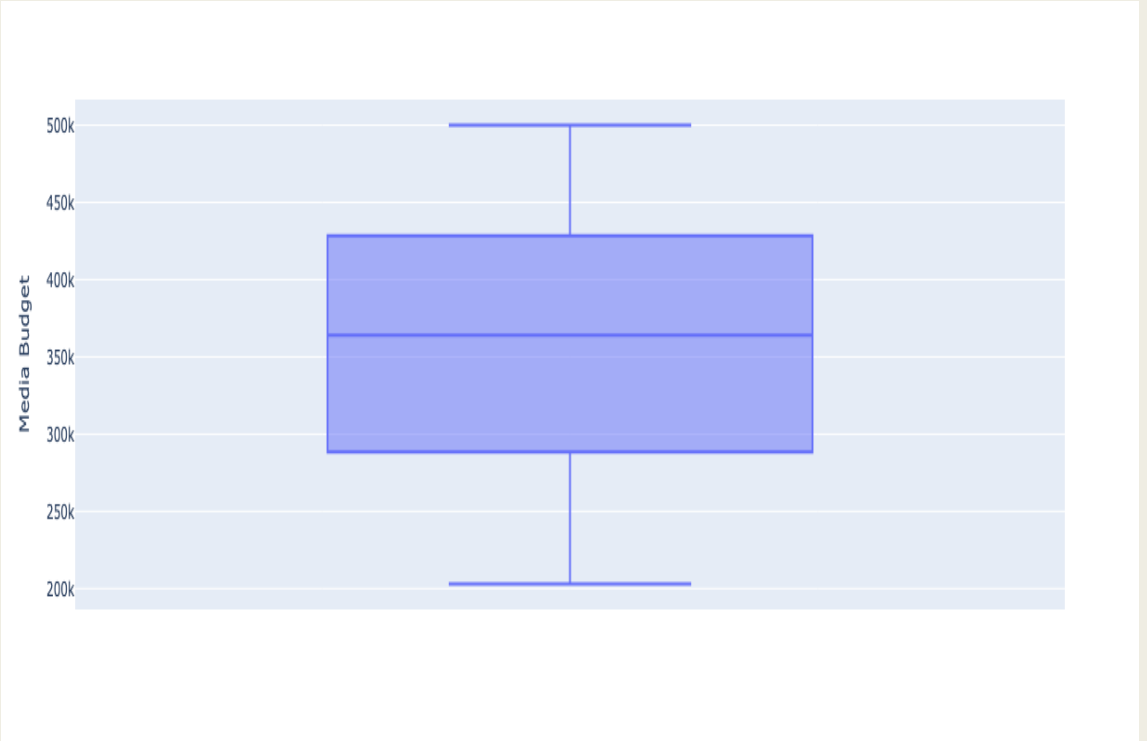


Brands

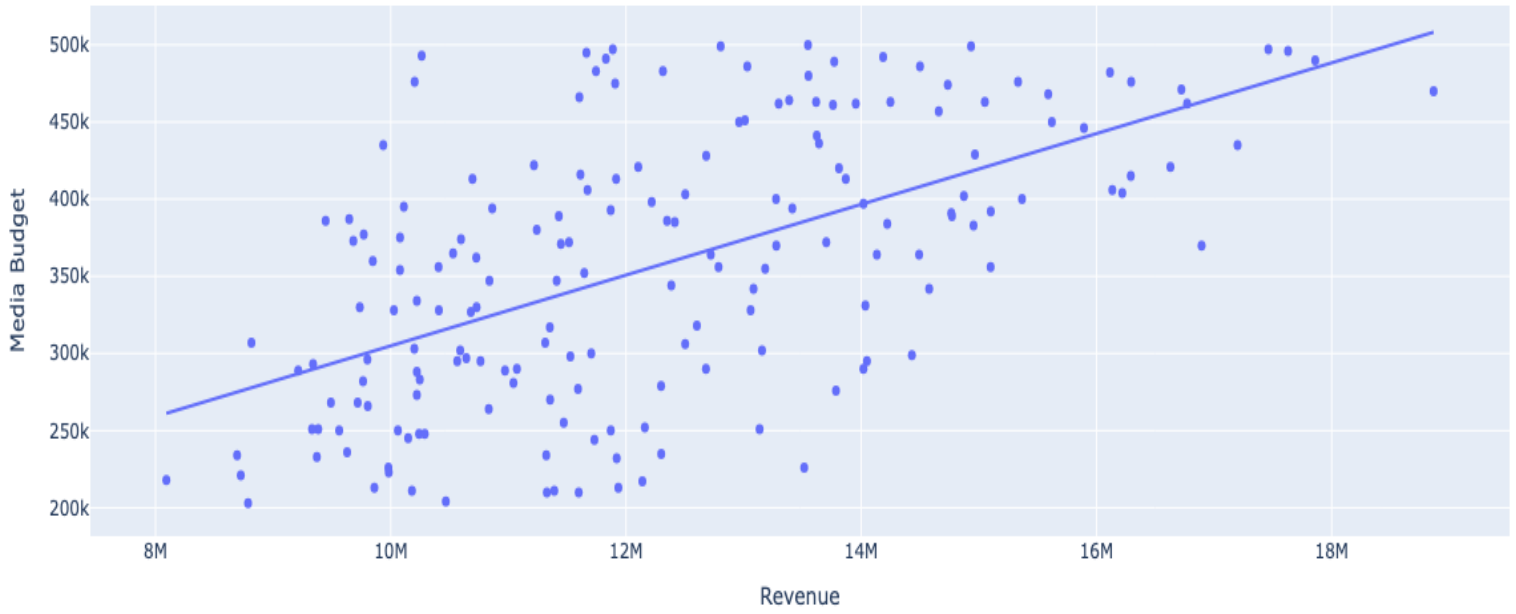
- There seems to be peaks during July and December, these patterns indicate a potential trend effect in the data due to the passage of time.
- Media budget shows that it might be cyclical in relation to brand allocation.

Outliers





Revenue shows potential outliers and an unbalanced distribution, whereas Media Budget is very balanced.

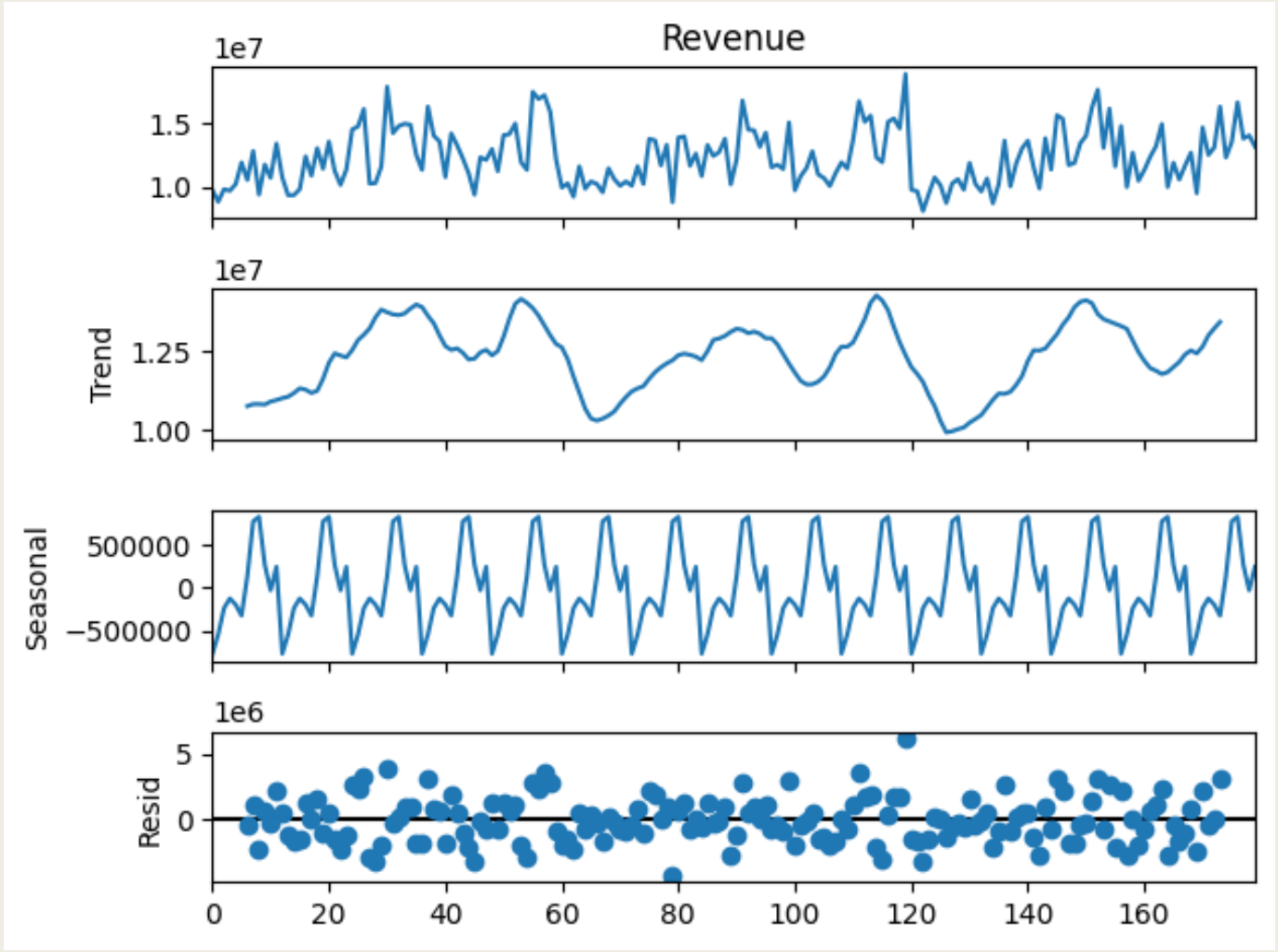


Revenue and Media Budget

- There is a positive relationship between Revenue and Media Budget, this supports the possibility that that the two can be included in a bivariate model.

Trend or no Trend?

- An additive seasonal decomposition of the revenue time series showed that a time series model is unlikely to provide a robust model, at least not without more feature engineering.



The Kendall-Tau correlation coefficients were also calculate to mathematically determine a trend component and none were significant at the 5% level.

Normality of Distribution

- Although the technique of testing the distribution of a sample for normality can be debated, in the spirit of thoroughness, three different techniques were used to determine whether the Revenue and Media Budget were normally distributed.

Test	Conclusion (Revenue and Media Budget)
Plotting mean and std	Inconclusive
Shapiro-Wilk test	Do not accept H0 that sample came from a population with normal distribution
QQ-plot	Visual does not indicate normal distribution

These tests all indicate that we should not assume that the data is normally distributed, therefore, moving forward we will use non-parametric tests that do not make assumptions about the distribution

Correlation Analysis: Kendal-Tau(Revenue and Media Budget)

- It is a non-parametric test therefore; it does not make assumptions about the distribution of data.
- It can also be used to calculate the trend in the data by setting itself as an ordinal representation of time. Kendall's Tau would represent the trend in the time series.

Assumptions:

- Variables tested are continuous
- Variables tested are monotonic (one consistent direction in relationship)

Note:

Tau correlation coefficient is often compared to Pearson correlation coefficient, however, its advantage in this context is that is also works when we are not sure if the relationship is ordinal or not.

Results:

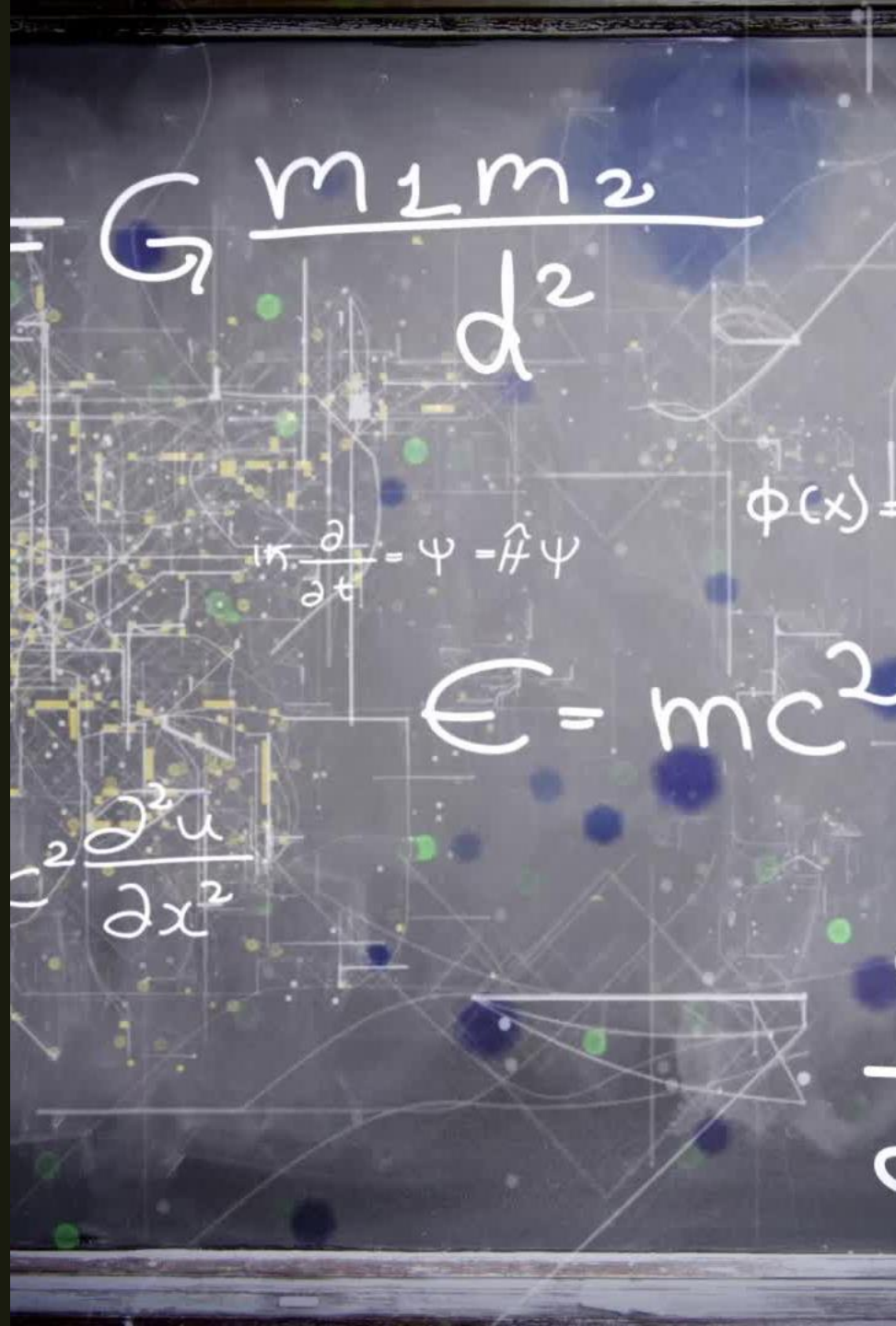
- Tau is calculated as 0.3996275611702023 7
- P-value is calculated as 1.85878001899808 3e-15

Conclusion

- Revenue and Media Budget are statistically positively correlated.

EDA Conclusion

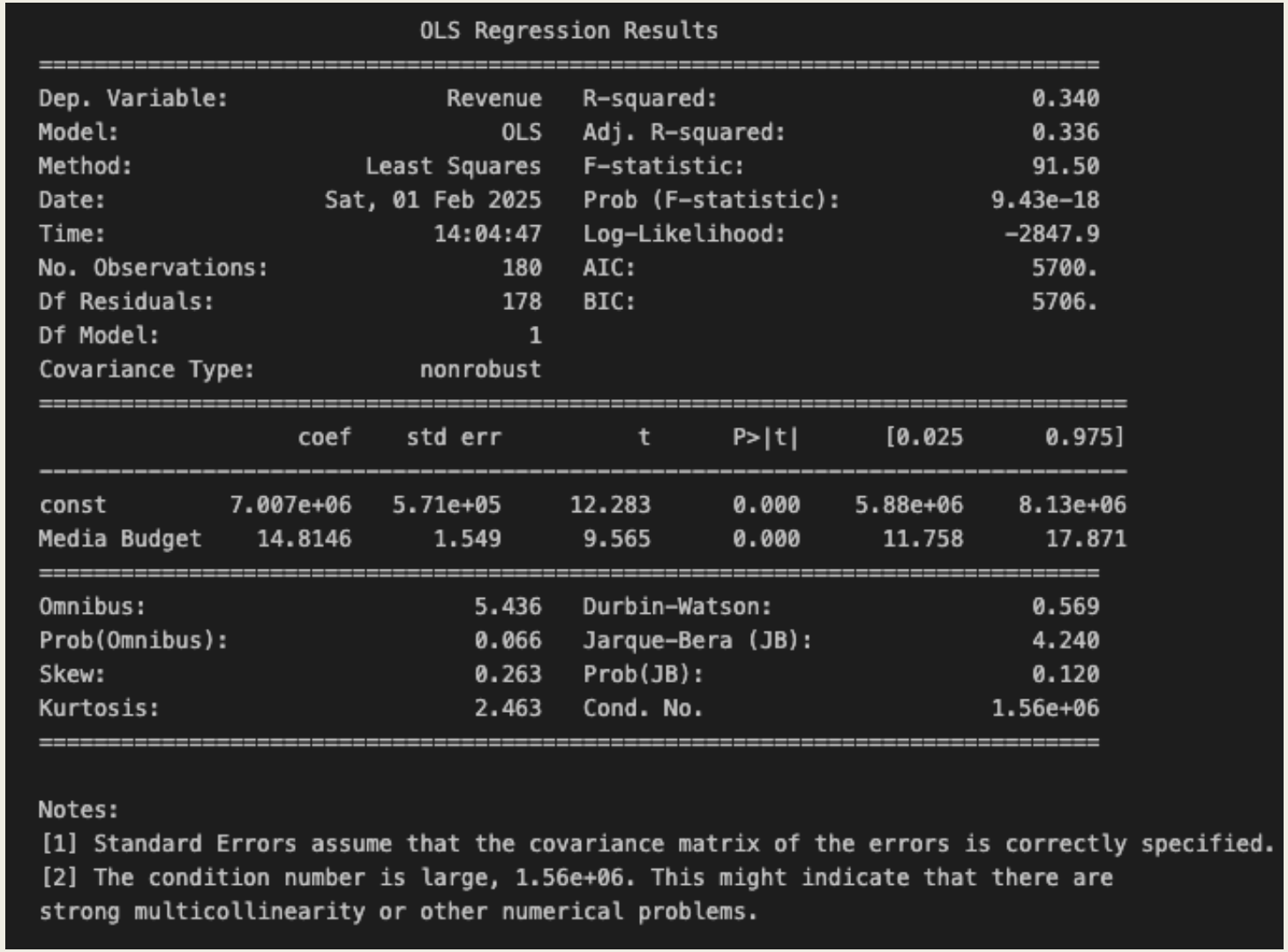
- Given the size of the dataset and the normality tests, we should not choose a model that is strictly dependent on a normal distribution.
- Therefore, going forward the tests used will likely be non-parametric and assume linearity.



MODELLING

Data Preparation Steps

1. Brand column was transferred by one-hot encoding. In the end this was not necessary as no amount of transformation resulted in a statistically significant Brand feature in the results of the model.
2. Normalisation using MixMax Scaling instead of scaling. Normalisation maintains the interpretability of the results and does not make assumptions about the distribution of the features and target variable.
3. OLS bivariate regression to determine size of influence of Budget Allocation on Revenue.
4. Then another linear regression was done with a Bayesian optimizer to allocate the total media budget.



OLS Regression Results

Dep. Variable:	Revenue	R-squared:	0.340
Model:	OLS	Adj. R-squared:	0.336
Method:	Least Squares	F-statistic:	91.50
Date:	Sat, 01 Feb 2025	Prob (F-statistic):	9.43e-18
Time:	14:04:47	Log-Likelihood:	-2847.9
No. Observations:	180	AIC:	5700.
Df Residuals:	178	BIC:	5706.
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	7.007e+06	5.71e+05	12.283	0.000	5.88e+06	8.13e+06
Media Budget	14.8146	1.549	9.565	0.000	11.758	17.871

Omnibus:	5.436	Durbin-Watson:	0.569
Prob(Omnibus):	0.066	Jarque-Bera (JB):	4.240
Skew:	0.263	Prob(JB):	0.120
Kurtosis:	2.463	Cond. No.	1.56e+06

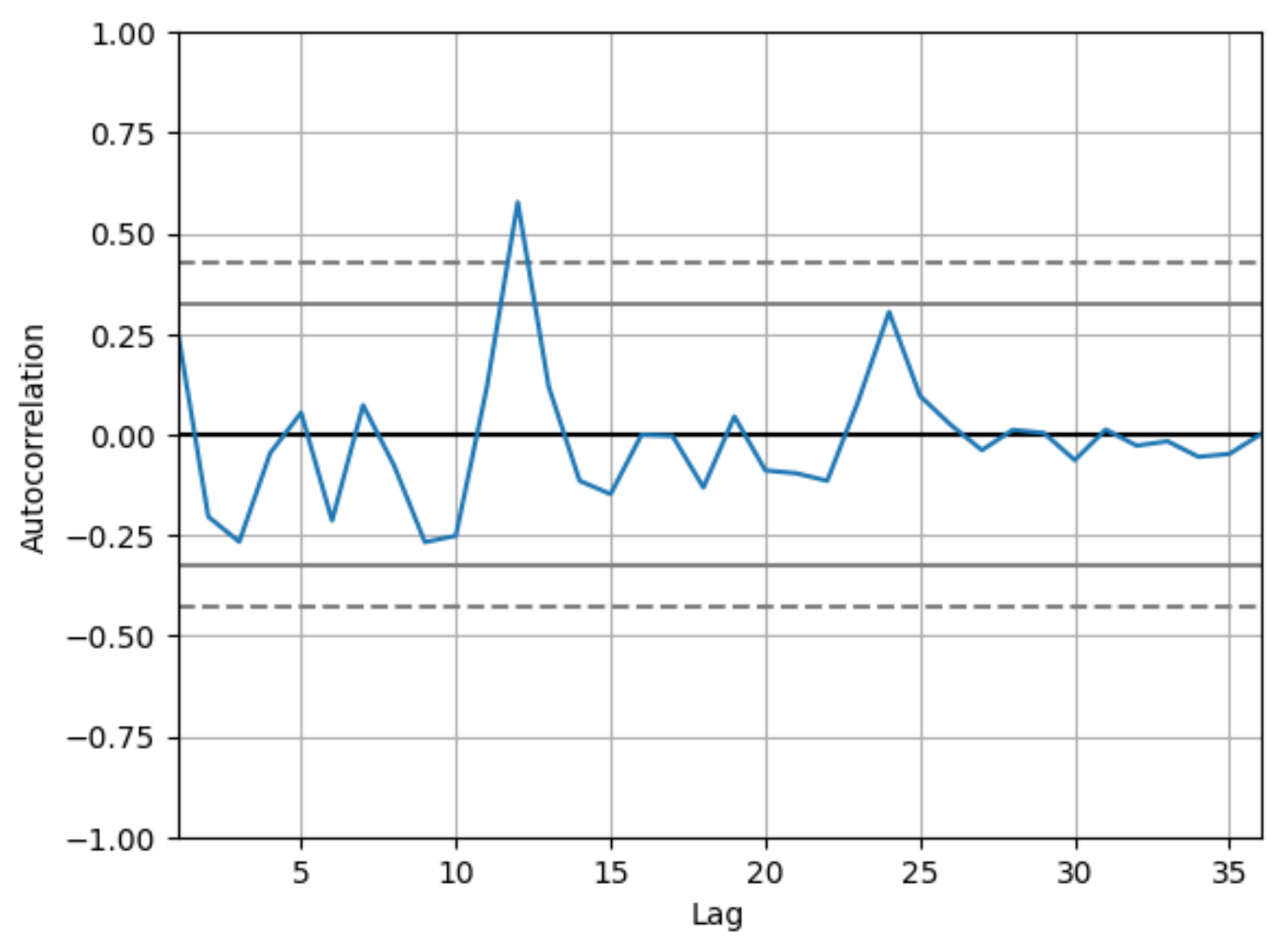
Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
[2] The condition number is large, 1.56e+06. This might indicate that there are strong multicollinearity or other numerical problems.

OLS Results

- Approximately 34% of the variation observed in the revenue is explained by the media budget. This is expected as the calculated strength of the correlation was significant but not strong.
- The variables were tested for autocorrelation, but none was found. However, due to the large variance observed in the EDA section this is not unexpected.

Budget Allocation 2023

- The results of the optimised budget allocation were exported and scaled to the previous years brand percentage allocation. Results and calculations are provided in the excel workbook '*feature_engineering*' on sheet '*Final Budget Allocation*'.
- Unfortunately, due to the lack of data the actual revenue predictions of this model were lower than expected. This is expected as we know from the OLS results that media budget alone cannot explain revenue.
- To remedy this situation a third model was created; an ARIMA model used to predict revenue based on it's lagged(past) values.



Revenue Predictions: ARIMA

- Looking at this autocorrelation plot we see that the lag only becomes significant at around lag 12. This implies that the lag effect is cyclical based on the previous year.

SARIMAX Results

Dep. Variable:	Client Market Size	No. Observations:	36
Model:	ARIMA(12, 1, 0)	Log Likelihood	-682.058
Date:	Sat, 01 Feb 2025	AIC	1390.116
Time:	15:26:06	BIC	1410.335
Sample:	01-01-2020	HQIC	1397.096
	- 12-01-2022		
Covariance Type:	opg		

	coef	std err	z	P>|z|	[0.025	0.975]
ar.L1	0.0227	0.031	0.739	0.460	-0.038	0.083
ar.L2	-0.1986	0.031	-6.507	0.000	-0.258	-0.139
ar.L3	-0.0572	0.028	-2.076	0.038	-0.111	-0.003
ar.L4	0.0377	0.039	0.977	0.329	-0.038	0.113
ar.L5	0.2954	0.024	12.503	0.000	0.249	0.342
ar.L6	-0.1484	0.026	-5.770	0.000	-0.199	-0.098
ar.L7	0.2166	0.016	13.260	0.000	0.185	0.249
ar.L8	-0.1096	0.038	-2.895	0.004	-0.184	-0.035
ar.L9	-0.1237	0.035	-3.577	0.000	-0.191	-0.056
ar.L10	-0.1520	0.019	-8.176	0.000	-0.188	-0.116
ar.L11	0.0745	0.029	2.552	0.011	0.017	0.132
ar.L12	-0.1464	0.012	-12.308	0.000	-0.170	-0.123
sigma2	1.107e+13	1.2e-15	9.23e+27	0.000	1.11e+13	1.11e+13

...

Warnings:
[1] Covariance matrix calculated using the outer product of gradients (complex-step).
[2] Covariance matrix is singular or near-singular, with condition number 2.53e+43. Standard errors may be unstable.

ARIMA Results

- As we can see the results are highly significant and implies that there is a lagged effect on revenue, and it is strong enough to model and predict.
- The revenue predictions of this model were exported and compared to the Linear regression revenue predictions.

BUDGET ALLOCATION

Date	Linear Model Revenue Predictions	ARIMA Predictions	Total Projected Budget Allocation	Return on ad Spend: Linear Model	Return on ad Spend: ARIMA
2023-01-01	57689047,2	73413031,61	2718042,479	21,2244833	27,00952328
2023-02-01	39726715,77	75848302,7	2546476,594	15,60066009	29,78558801
2023-03-01	19800702,49	47675880,72	965959,2404	20,49848654	49,35599633
2023-04-01	42880500,33	58040348,5	2266116,131	18,92246375	25,6122569
2023-05-01	53978913,94	54542927,32	2525686,195	21,37197964	21,59529059
2023-06-01	45129986,23	53666579,44	1290104,089	34,9816628	41,598643
2023-07-01	37523951,78	53122289,04	1688221,607	22,22691123	31,46641936
2023-08-01	19901547,95	67473191,99	1491846,893	13,34020806	45,22796025
2023-09-01	21402577,42	70074170,95	2331438,674	9,179987302	30,05619309
2023-10-01	9898262,166	61026341,25	2055671,515	4,815099151	29,68681562
2023-11-01	9122917,924	72513160,37	396136,1313	23,02975468	183,0511146
2023-12-01	54287494,61	59443653,08	2224300,453	24,40654748	26,72465089
	€411 342 617,81	€746 839 876,99	€22 500 000,00	19,133187	45,09753766

Budget Allocation: Results

- As we can see, the ARIMA model projections are inline with historical data, if not slightly better.
- We have also increased the Return on spend/ROI; predicted to be 45,1 versus the 36,44 from 2022. Now each euro allocated to the media budget will generate an extra €10.

Budget Allocation: Results Continued

With the Bayesian optimization we have managed to increase projected revenue by 2%, a major improvement from the previous years which had negative growth. 2% Does not sound like a lot initially, but it represents a growth of € 17,307,469.99.

Year	Total Revenue	Growth
2020	€749 179 945,00	0%
2021	€737 035 777,00	-2%
2022	€729 532 407,00	-1%
2023	€746 839 876,99	2%

THANK YOU!

For extra information and proofs, please check the notebooks provided or shoot me an email.