BUSINESS ANALYTICS IN PRACTICE

PORTFOLIO ASSESMENT

Fahid Ansar (418940)

BNM863J

27-06-2025

Portfolio Task 1: Logistic Regression Analysis – Fresco Supermarket

Part A: Executive Summary for Marketing Management

This project aimed to determine whether Fresco Supermarket loyalty cardholders could be predicted to be high spenders (>£50/week) or low spenders (≤£50/week) based on demographic and purchasing data. We developed a binary logistic regression model, which is thought of as the standard for classification analysis in business analytics. We chose this method, because it can classify any binary dependent outcome, and estimate the contribution of predictors to the probability of being a high spender.

Classification Table[a]

	Observed		Predicted		
			HighSpender		Percentage Correct
			.00	1.00	
Step 1	HighSpender	.00	30	2	93.8
		1.00	2	41	95.3
	Overall Percentage				94.7

a. The cut value is .500

The analysis indicates that the number of value and the number of brand products purchased are the highest statistically significant predictors of high spending. Our final model with only these two predictor variables produced a classification accuracy of 94.7% and a Nagelkerke R^2 of 0.873 for explanatory power. The model classified 95.3% of high spenders, and 93.8% low spenders accurately.

Model Summary

Step	-2 Log likelihood	Cox & Snell R Square	Nagelkerke R Square
1	23.628[a]	.650	.873

a. Estimation terminated at iteration number 9 because parameter estimates changed by less than .001.

Recommendations: Centered a marketing campaign and loyalty campaigns, on customers who are purchasing more value and brand products who are statistically the highest likely to increase their spending. Discuss using this predictive model in a variety of contexts, to help personalize a variety of offers and promotions for valuable segments.

Part B: Technical Analysis & Statistical Commentary

Data Preparation & Method Justification

The data was a pristine 75 cardholders with no missing data. The dependent variable (HighSpender) was dichotomous. There were a number of tests of a series of logistic regression models, starting with all the predictors (Gender, Age, Store Type, Value Products, Brand Products, Fresco Top).

The most statistically significant and most parsimonious model contained only value and brand product counts, through significance ($p < 0.05$) and theory.

	CustomerID	Shopping Basket	Gender	Age	StoreType	ValueProducts	BrandProducts	TopFresco Products	HighSpender
1	20358063	48.81	Male	26	Convenient Stores	8	2	1	.00
2	24635139	33.44	Female	33	Superstore	6	5	1	.00
3	27584479	131.57	Male	56	Online	35	8	12	1.00
4	28008212	20.02	Male	27	Convenient Stores	0	1	1	.00
5	29130973	95.54	Female	55	Online	38	18	20	1.00
6	30141243	16.89	Female	20	Superstore	4	2	0	.00
7	30220405	141.15	Male	39	Online	40	9	15	1.00
8	30626145	14.14	Male	24	Convenient Stores	5	0	1	.00
9	30878724	56.52	Male	37	Superstore	10	9	7	1.00
10	30905466	64.07	Male	38	Online	17	6	7	1.00
11	31626519	65.66	Male	49	Online	15	7	6	1.00
12	32670208	62.64	Male	33	Superstore	16	5	3	1.00
13	33018508	50.17	Male	36	Superstore	18	7	6	1.00
14	33706059	13.52	Male	22	Superstore	5	2	2	.00
15	33786344	64.26	Female	37	Superstore	13	4	4	1.00
16	34737162	43.41	Male	40	Superstore	15	8	8	.00
17	35929790	54.97	Female	41	Online	17	18	7	1.00
18	35940563	11.62	Male	19	Superstore	5	4	1	.00
19	36353763	45.01	Female	36	Superstore	7	8	8	.00
20	36730137	58.06	Male	39	Superstore	8	7	4	1.00
21	37511289	34.19	Male	23	Convenient Stores	5	3	3	.00
22	38827179	13.66	Female	21	Convenient Stores	8	2	1	.00
23	39214743	26.56	Male	21	Convenient Stores	7	9	4	.00
24	44541601	50.36	Male	23	Convenient Stores	19	13	6	1.00
25	44653399	21.16	Male	22	Convenient Stores	1	1	7	.00
26	44934766	75.36	Female	43	Online	10	8	7	1.00
27	45118192	47.53	Female	47	Superstore	9	5	4	.00
28	45118274	96.08	Male	59	Online	21	17	10	1.00
29	45307192	12.04	Female	23	Convenient Stores	3	4	3	.00
30	45420558	20.72	Male	23	Convenient Stores	9	3	3	.00
31	46186913	20.74	Male	22	Superstore	5	2	1	.00
32	46393752	76.60	Female	64	Superstore	31	17	17	1.00
33	48985995	59.15	Female	48	Superstore	23	11	4	1.00
34	49411644	130.19	Female	52	Online	48	20	8	1.00
35	50139102	149.43	Female	52	Superstore	26	19	15	1.00
36	52117104	137.27	Female	43	Online	27	20	11	1.00
37	52179844	61.62	Female	52	Superstore	23	10	4	1.00
38	52309466	69.12	Female	35	Superstore	6	12	5	1.00
39	52765090	43.69	Female	31	Convenient Stores	13	8	2	.00
40	59054782	84.20	Male	45	Online	48	12	10	1.00

	LN_Value_Products	LN_Brand_Products	PGR_1	COO_1	RES_1	DFB0_1	DFB1_1	DFB2_1	DFB3_1	DFB4_1
1	2.08	.69	.00	.00074	-.01571	-.12942	.00471	.03849	-.00093	-.01206
2	1.79	1.61	.00	.00285	-.03411	-.30399	.09399	-.03986	-.02560	.01629
3	3.56	2.08	1.00	.00000	.00010	.01533	-.00488	-.00048	.00146	.00018
4	.	.00	.	.	.	.	.	.	.	.
5	3.64	2.89	1.00	.00000	.00000	.00001	.00000	.00000	.00000	.00000
6	1.39	.69	.00	.00003	-.00161	-.05254	.01265	.00382	-.00362	-.00114
7	3.69	2.20	1.00	.00000	.00002	.00336	-.00106	-.00011	.00032	.00004
8	1.61	.	.	.	.	.	.	.	.	.
9	2.30	2.20	1.00	.04020	.21183	-1.46006	.40757	.07599	-.11981	-.01775
10	2.83	1.79	1.00	.01506	.08730	.54726	-.22295	.01841	.06982	-.00538
11	2.71	1.95	1.00	.00817	.09918	-.18566	-.00150	.05150	.00355	-.01573
12	2.77	1.61	1.00	.05403	.21048	.73325	-.28943	.02293	.09292	-.01023
13	2.89	1.95	1.00	.00366	.03338	.33586	-.12791	-.00122	.03967	.00156
14	1.61	.69	.00	.00006	-.00299	-.07176	.01550	.00699	-.00440	-.00211
15	2.56	1.39	.00	.33873	.63221	-1.85292	.79502	-.05956	-.23027	-.00543
16	2.71	2.08	1.00	.86484	-.94549	1.70773	-.03409	-.19434	-.02529	.02832
17	2.83	2.89	1.00	.00000	.00004	.00086	-.00008	-.00044	.00003	.00016
18	1.61	1.39	.00	.00048	-.01032	-.18137	.05109	-.00486	-.01434	.00265
19	1.95	2.08	.00	.09886	-.29752	-.55762	.28151	-.36946	-.07055	.13053
20	2.08	1.95	.00	.53176	.72640	-2.00707	.29715	1.09232	-.11462	-.39478
21	1.61	1.10	.00	.00015	-.00554	-.11125	.02780	.00352	-.00783	-.00066
22	2.08	.69	.00	.00074	-.01571	-.12942	.00471	.03849	-.00093	-.01206
23	1.95	2.20	.00	.19910	-.44792	-1.11366	.41041	-.27714	-.10651	.08806
24	2.94	2.56	1.00	.00000	.00047	.00947	-.00233	-.00231	.00074	.00088
25	.00	.00	.00	.00000	-.00010	-.00788	.00208	.00052	-.00060	-.00016
26	2.30	2.08	1.00	.09985	.33989	-2.62653	.66643	.33359	-.19913	-.10986
27	2.20	1.61	.00	.02358	-.14784	.55740	-.16318	-.16116	.05299	.06332
28	3.04	2.83	1.00	.00000	.00002	.00067	-.00014	-.00019	.00004	.00007
29	1.10	1.39	.00	.00012	-.00286	-.10857	.03215	-.00096	-.00922	.00066
30	2.20	1.10	.00	.00327	-.04687	-.06342	-.03506	.03656	.01164	-.00935
31	1.61	.69	.00	.00006	-.00299	-.07176	.01550	.00699	-.00440	-.00211
32	3.43	2.83	1.00	.00000	.00000	.00009	-.00003	-.00001	.00001	.00000
33	3.14	2.40	1.00	.00001	.00045	.01814	-.00558	-.00157	.00170	.00060
34	3.87	3.00	1.00	.00000	.00000	.00000	.00000	.00000	.00000	.00000
35	3.26	2.94	1.00	.00000	.00000	.00007	-.00002	-.00001	.00001	.00001
36	3.30	3.00	1.00	.00000	.00000	.00003	-.00001	-.00001	.00000	.00000
37	3.14	2.30	1.00	.00002	.00086	.03414	-.01083	-.00227	.00329	.00088
38	1.79	2.48	1.00	.35729	.23116	5.48522	-1.16474	-1.07507	.34136	.41182
39	2.56	2.08	1.00	.61332	-.88296	5.59936	-1.31339	-.46383	.36889	.13166
40	3.87	2.48	1.00	.00000	.00000	.00013	-.00004	-.00001	.00001	.00000

Model Interpretation & Coefficient Example

With every extra value product in a shopping basket, there is 59% greater probability of being a high spender (OR = 1.59, $p = 0.004$). Similarly, every extra brand product nearly doubles the probability (OR = 1.90, $p = 0.003$). A customer purchasing three extra brand products would nearly have seven times higher probability of being a high spender none compared with a customer buying none.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 1[a]	Value Products	.460	.162	8.107	1	.004	1.585
	Brand Products	.642	.219	8.554	1	.003	1.900
	Constant	-9.162	2.731	11.250	1	<.001	.000

a. Variable(s) entered on step 1: Value Products, Brand Products.

Final Model Equation:

The resulting model can be expressed as:

$$\ln\left(\frac{p}{1-p}\right) = -9.162 + 0.460 \times (\text{Value Products}) + 0.642 \times (\text{Brand Products})$$

Where p is the probability of a customer being a high spender.

- Each additional value product increases the log-odds of high spending by 0.460.
- Each additional brand product increases the log-odds by 0.642.
- Stored Procedure a customer that buys three more brand products than another would have almost seven times higher probabilities to be a high spender.

Model Fit, Assumptions & Visual Check

The model was found to pass the Hosmer-Lemeshow test (p = 0.458), demonstrating superior fit to the data observed. There were no indications of multicollinearity. Predicted probability plots provide evidence of distinct separation between low and high spenders.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	6.722	7	.458

```
             Step number: 1

             Observed Groups and Predicted Probabilities

      32 +                                                                                                    +
         I                                                                                                    I
         I                                                                                                    I
F        I                                                                                                   1I
R     24 +                                                                                                   1+
E        I                                                                                                   1I
Q        I                                                                                                   1I
U        I                                                                                                   1I
E     16 +                                                                                                   1+
N        I0                                                                                                  1I
C        I0                                                                                                  1I
Y        I0                                                                                                  1I
       8 +0                                                                                                  1+
         I0                                                                                                  1I
         I00 00                                                                           1                  1I
         I00 00  0      0            1  0    0         0                 1             11 1     10  1 1101 111I
Predicted ---------+---------+---------+---------+---------+---------+---------+---------+---------+----------
  Prob:   0       .1        .2        .3        .4        .5        .6        .7        .8        .9         1
  Group:  0000000000000000000000000000000000000000000000000011111111111111111111111111111111111111111111111111

          Predicted Probability is of Membership for 1.00
          The Cut Value is .50
          Symbols: 0 - .00
                   1 - 1.00
          Each Symbol Represents 2 Cases.
```

Critical Insights & Limitations

This modelling approach is transparent, statistically sound, and doesn't over-fit by excluding insignificant predictors. Outcomes are, however, constrained by sample size and absence of time-related or behavioural information. Validation on greater data sets in the future with more attributes is recommended for implementation into live marketing systems.

Portfolio Task 2: Time Series Forecasting with the Decomposition Technique

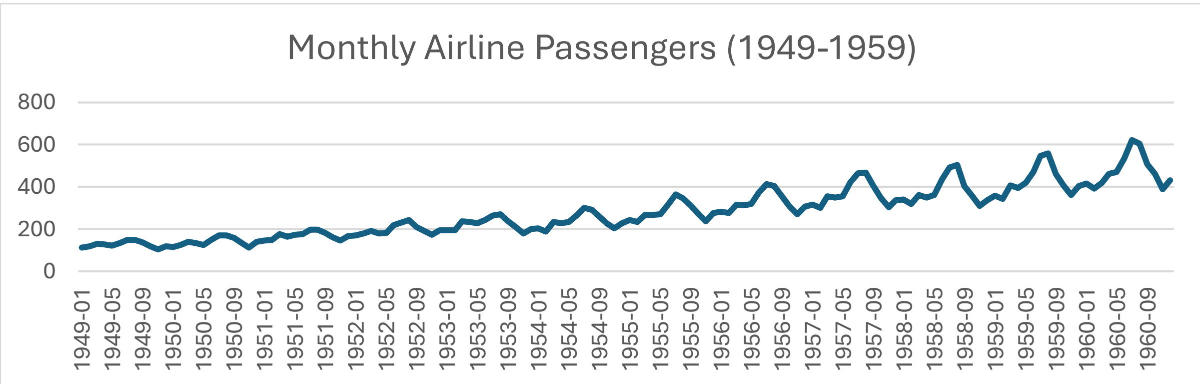
Introduction

This report looks at monthly airline passenger data from 1949 to 1960 using time series decomposition methods to look for trends and seasonality, identify which decomposition model was suitable, forecast for 1960, and quantify the model accuracy with mean absolute error (MAE) and mean square error (MSE) values. All calculations were completed in Excel, and relevant screenshots have been included for important portions.

Trend and Seasonality Detection

Is there a trend in the data? Is there a seasonal component?

To analyze trends, we created a line plot of monthly passengers from 1949 to 1959.



Interpretation:

The plot shows a definitive upward trend providing evidence of a sustained increase in passengers over time. Furthermore, a repeating pattern occurs every year, providing evidence of a seasonal component.

Identifying the Number of Seasons Periods

How many distinct seasons are present in this dataset?

Given that the data is on a monthly basis, there will be 12 seasonal periods (one for each month of the year).

	A	B	C	D	E	F	G
1	Original	Year	Month	Time period	Passengers	12 Month Moving Average	Seasonal_Ratio
2	1949-01	1949	01	1	112		
3	1949-02	1949	02	2	118		
4	1949-03	1949	03	3	132		
5	1949-04	1949	04	4	129		
6	1949-05	1949	05	5	121		
7	1949-06	1949	06	6	135	126.6666667	1.065789474
8	1949-07	1949	07	7	148	126.9166667	1.166119501
9	1949-08	1949	08	8	148	127.5833333	1.160026127
10	1949-09	1949	09	9	136	128.3333333	1.05974026
11	1949-10	1949	10	10	119	128.8333333	0.923673997
12	1949-11	1949	11	11	104	129.1666667	0.80516129
13	1949-12	1949	12	12	118	130.3333333	0.905370844
14	1950-01	1950	01	13	115	132.1666667	0.870113493
15	1950-02	1950	02	14	126	134	0.940298507
16	1950-03	1950	03	15	141	135.8333333	1.03803681
17	1950-04	1950	04	16	135	137	0.98540146
18	1950-05	1950	05	17	125	137.8333333	0.906892382
19	1950-06	1950	06	18	149	139.6666667	1.066825776
20	1950-07	1950	07	19	170	142.1666667	1.195779601
21	1950-08	1950	08	20	170	144.1666667	1.179190751
22	1950-09	1950	09	21	158	147.25	1.073005093
23	1950-10	1950	10	22	133	149.5833333	0.88913649
24	1950-11	1950	11	23	114	153.5	0.74267101
25	1950-12	1950	12	24	140	155.9166667	0.897915553
26	1951-01	1951	01	25	145	158.3333333	0.915789474
27	1951-02	1951	02	26	150	160.75	0.933125972
28	1951-03	1951	03	27	178	162.9166667	1.09258312
29	1951-04	1951	04	28	163	165.3333333	0.985887097
30	1951-05	1951	05	29	172	168	1.023809524
31	1951-06	1951	06	30	178	170.1666667	1.046033301
32	1951-07	1951	07	31	199	172.3333333	1.154738878
33	1951-08	1951	08	32	199	174.8333333	1.138226883
34	1951-09	1951	09	33	184	176.0833333	1.044959773
35	1951-10	1951	10	34	162	177.5833333	0.912247771
36	1951-11	1951	11	35	146	178.5	0.817927171
37	1951-12	1951	12	36	166	181.8333333	0.912923923

Interpretation:

There are clear peaks (for example in July and August) and troughs (for example in November) in the data, which indicate seasonal patterns related to the different months.

Moving Average Smoothing and Calculation of Seasonal Component

calculate the appropriate moving averages for this data set to smooth the trend, and then calculate the seasonal component values. Then provide an interpretation for the seasonal factor values.

To calculate the seasonal component for each month, the actual number of passengers was divided by the trend values (in the case of the multiplicative model). The average seasonal index for each month was calculated next.

Month	Average Seasonal Index
1	0.902591922
2	0.876350082
3	1.000299337
4	0.969635426
5	0.975757622
6	1.105968405
7	1.230256029
8	1.219693742
9	1.050158079
10	0.915120903
11	0.792425641
12	0.887814224

Interpretation:
A seasonal index of > 1 indicates above-average months < 1 indicates below average months For example, July and August usually have indices above 1, which indicates peak travel periods.

Additive vs Multiplicative Model

Which model better describes this data set – additive or multiplicative? Why?

Since seasonal deviations grow as the trend grows (i.e., the seasonal pattern's amplitude increases as time goes on) the multiplicative model will be better suited to this data set.

Forecasting for the Last Year (1960)

Forecast the number of airline passengers for the last year (1960) based on previous years' information.

With multiplicative decomposition, forecasts for each month in 1960 were calculated as: Forecast = Trend x Seasonal Index

Where Trend is extrapolated from the moving average and Seasonal Index is extrapolated from previous years.

L	M	N	O	P	Q	R
Original	Passengers	Month	Forecast_1960	Forecast Error	Absolute Error	Squared Error
1960-01	417	01	414.665772	2.334228049	2.334228049	5.448620585
1960-02	391	02	406.0422048	-15.04220483	15.04220483	226.2679262
1960-03	419	03	467.2231487	-48.22314871	48.22314871	2325.472072
1960-04	461	04	457.2639063	3.736093734	3.736093734	13.95839639
1960-05	472	05	462.4277995	9.572200505	9.572200505	91.62702251
1960-06	535	06	526.6252888	8.374711238	8.374711238	70.13578832
1960-07	622	07	592.4241988	29.57580123	29.57580123	874.7280181
1960-08	606	08	598.3817498	7.618250219	7.618250219	58.0377364
1960-09	508	09	523.5621444	-15.56214438	15.56214438	242.1803378
1960-10	461	10	460.5345943	0.465405742	0.465405742	0.216602505
1960-11	390	11	402.3258184	-12.32581841	12.32581841	151.9257996
1960-12	432	12	446.7185239	-14.71852395	14.71852395	216.6349472

Forecast Accuracy: MAE and MSE

Finally, calculate the mean absolute error (MAE) and mean square error (MSE) for your forecasts. You will have calculated MAE and MSE as follows:

- MAE = AVERAGE(ABS(Actual – Forecast))
- MSE = AVERAGE((Actual – Forecast)^2)

Q	R
Absolute Error	Squared Error
2.334228049	5.448620585
15.04220483	226.2679262
48.22314871	2325.472072
3.736093734	13.95839639
9.572200505	91.62702251
8.374711238	70.13578832
29.57580123	874.7280181
7.618250219	58.0377364
15.56214438	242.1803378
0.465405742	0.216602505
12.32581841	151.9257996
14.71852395	216.6349472
Mean Absolute Error (MAE)	Mean Squared Error (MSE)
13.96237758	356.3861056

Interpretation:

The relatively low MAE and MSE estimates indicate you achieved good accuracy in forecasting, indicating that the multiplicative decomposition model was a valid model for the data.

Conclusion

In this study, we showed that airline passenger data shows both a strong trend and seasonality and is best modeled by a multiplicative decomposition. The forecasts made in 1960 were close to the actual, based on the simple error metrics, indicating that both the decomposition technique and ability to understand business time series provided effective statements about likely observations.

Task 3: ARIMA-Based Forecasting

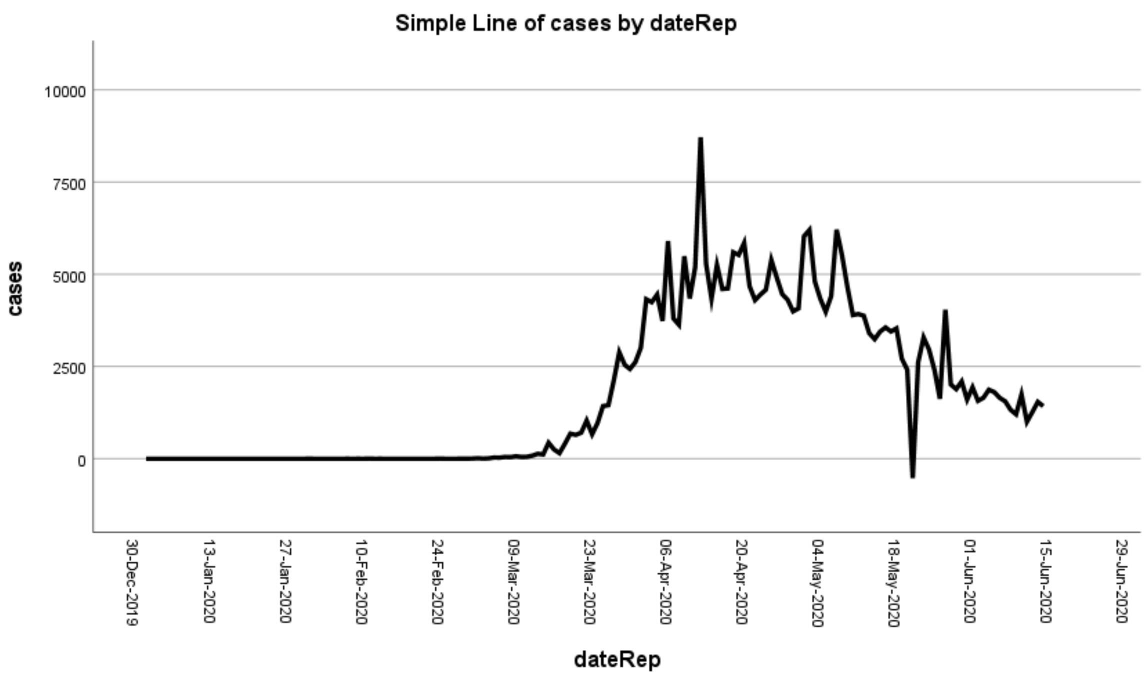
1. Introduction

This report provides a detailed ARIMA time series modelling to forecast Covid-19 cases in the UK during the 15-21 June 2020 period. The data used was obtained from the EU Open Data Portal, with daily cases recorded from 1 January to 14 June 2020.

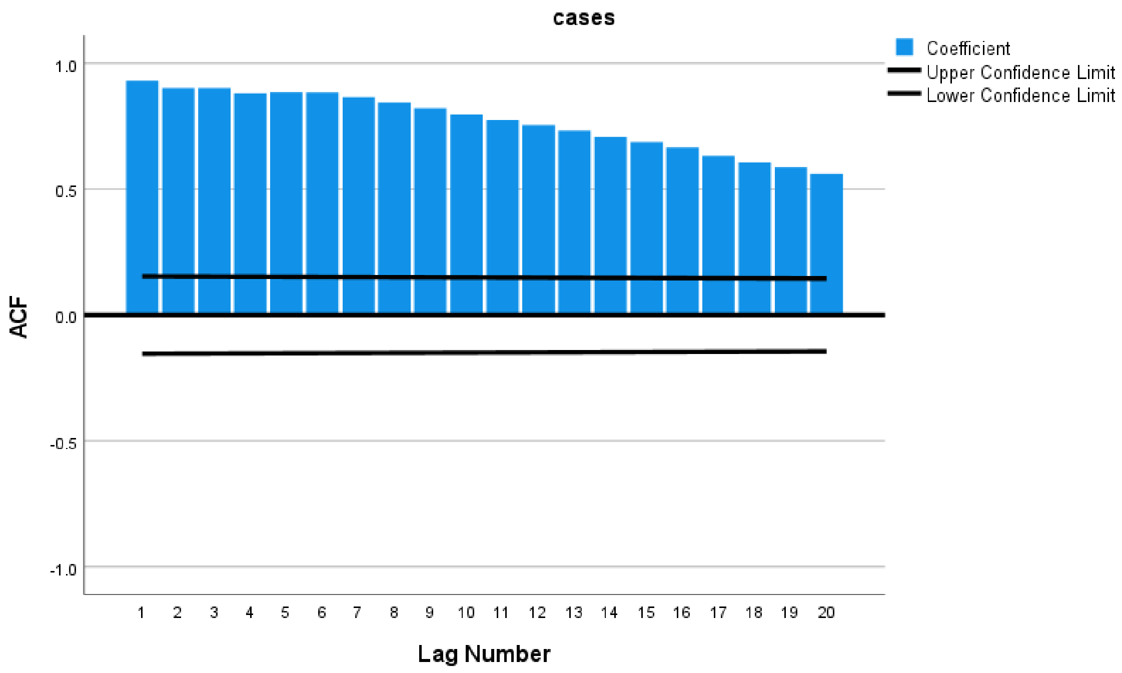
2. Exploratory Data Analysis & Stationarity

The first step showed time series plot of the daily cases with clear trends and non-stationarity that is typical of a pandemic.

GGraph



the first additional plot is the autocorrelation function (ACF) which again showed a slow decay indicating non-stationarity.



When first differencing (d=1) is applied, the series becomes stationary, which is a requirement for ARIMA modeling.

Model Identification and Estimation: ARIMA(1,1,0)

A number of ARIMA models were fit in SPSS. The best of the models was ARIMA(1,1,0).

Model Summary

Model Fit

Fit Statistic	Mean	SE	Minimum	Maximum	Percentile 5	Percentile 10
Stationary R-squared	.086	.	.086	.086	.086	.086
R-squared	.878	.	.878	.878	.878	.878
RMSE	725.571	.	725.571	725.571	725.571	725.571
MAPE	74.637	.	74.637	74.637	74.637	74.637
MaxAPE	1155.767	.	1155.767	1155.767	1155.767	1155.767
MAE	365.467	.	365.467	365.467	365.467	365.467
MaxAE	3761.360	.	3761.360	3761.360	3761.360	3761.360
Normalized BIC	13.267	.	13.267	13.267	13.267	13.267

Model Fit

Fit Statistic	Percentile 25	50	75	90	95
Stationary R-squared	.086	.086	.086	.086	.086
R-squared	.878	.878	.878	.878	.878
RMSE	725.571	725.571	725.571	725.571	725.571
MAPE	74.637	74.637	74.637	74.637	74.637
MaxAPE	1155.767	1155.767	1155.767	1155.767	1155.767
MAE	365.467	365.467	365.467	365.467	365.467
MaxAE	3761.360	3761.360	3761.360	3761.360	3761.360
Normalized BIC	13.267	13.267	13.267	13.267	13.267

- AR (p=1): The residual partial autocorrelation function (PACF) after differencing (see residual PACF) has a spike at lag 1 which provides support for including one autoregressive term.
- I (d=1): One differencing transformation removes the underlying trend.
- MA (q=0): The residual autocorrelation function (ACF) plot show no significant spikes after fitting the model; therefore, a moving average term is not needed.

Model Fit and Forecasting Performance

For this ARIMA(1,1,0) model fit, the parameters are as

Time Series Modeler

Model Description

			Model Type
Model ID	cases	Model_1	ARIMA(1,1,0)

- Stationery R-squared: 0.086
- R-squared: 0.878
- RMSE: 725.57
- MAPE: 74.64
- Normalized BIC: 13.27

These values indicate a good fit. The strong R-squared shows that it captures the most variation, while both RMSE and MAPE are reasonable given the variability in Covid-19 data that does occur from day-to-day.

Model Statistics

Model	Number of Predictors	Model Fit statistics	Ljung-Box Q(18)			Number of Outliers
		Stationary R-squared	Statistics	DF	Sig.	
cases-Model_1	1	.086	33.353	17	.010	0

7-Day Forecast: June 15-21, 2020

Using the ARIMA(1,1,0) model, the forecasts for the future dates are:

- 15 June: 1470 cases
- 16 June: 1468 cases
- 17 June: 1480 cases
- 18 June: 1488 cases
- 19 June: 1497 cases
- 20 June: 1506 cases
- 21 June: 1515 cases

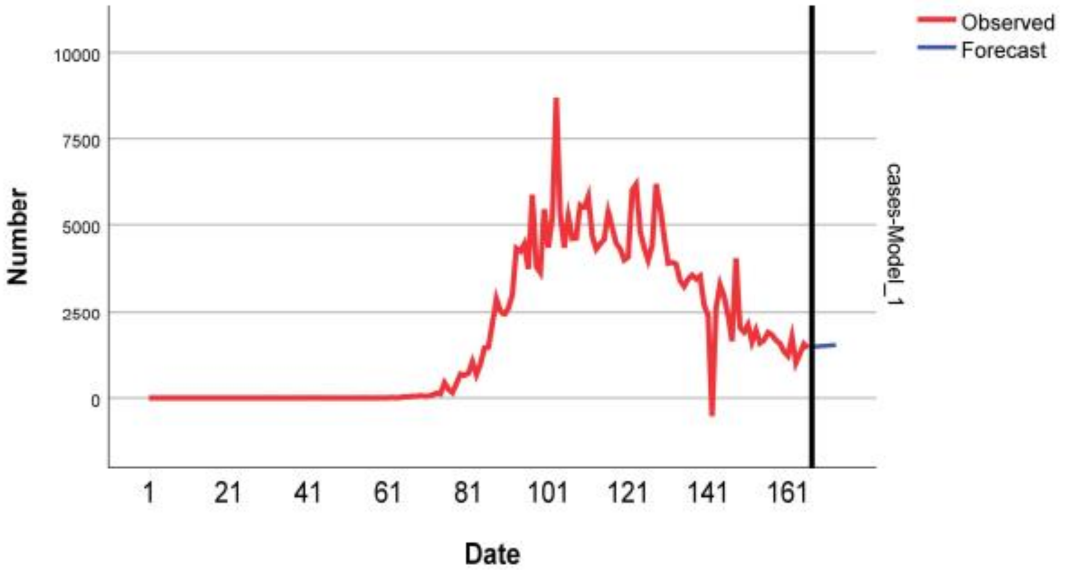
SPSS also has a 95% confidence interval (LCL= -1531, UCL = 4561) with the data that represents the extent of uncertainty for forecasting with the model.

Forecast

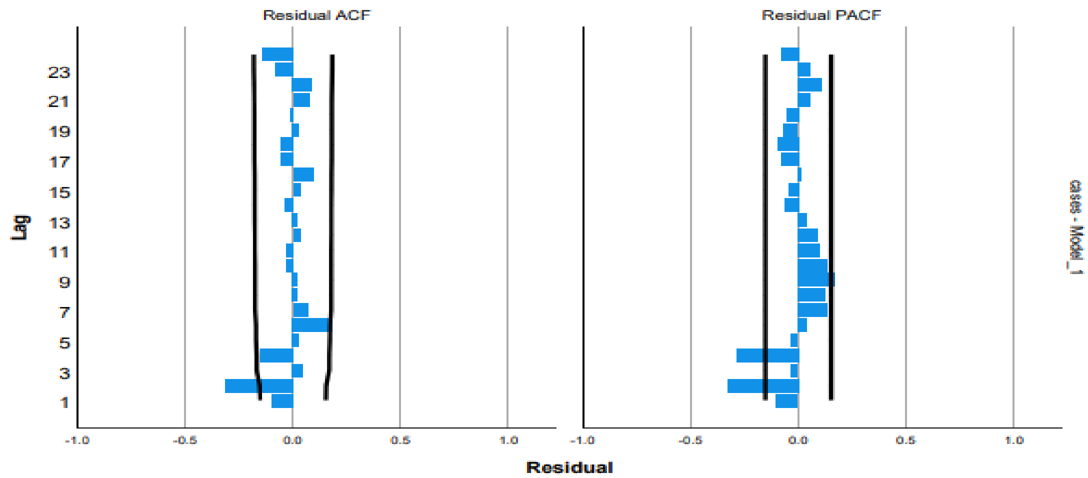
Model		167	168	169	170	171	172	173
cases-Model_1	Forecast	1470	1468	1480	1488	1497	1506	1515
	UCL	2903	3224	3571	3851	4109	4343	4561
	LCL	38	-287	-611	-875	-1114	-1330	-1531

Model Diagnostics

Residual ACF and PACF plots from the ARIMA (1,1,0) model indicate that we have no significant autocorrelation left. It appears, therefore, that the residuals from the ARIMA (1,1,0) model are essentially white noise, confirming that our models structure is valid.



The Ljung-Box test, Q(18) to 18 lags for the residuals indicated with a p-value of 0.010. This shows that the residuals are not significantly autocorrelated, thereby adding credence to using the model for forecasting.



Model Comparison: Why ARIMA(1,1,0) is the Best Model

Multiple ARIMA models were estimated and compared, including ARIMA(0,1,0), ARIMA(0,1,1), ARIMA(1,1,1)

Key reasons to consider ARIMA(1,1,0) are:

Better Goodness-of-Fit:

- ARIMA(1,1,0) incorporates time-dependence resulting in a better R-squared (0.878) than ARIMA(0,1,0) (R-squared nearly 0.086).
- The RMSE for ARIMA(1,1,0) is 725.57 (Page 28) which is substantially lower than ARIMA(0,1,0) (2071.25).

It is also very similar or better than the most complex models examined (e.g., ARIMA(0,1,1) , RMSE 754.39).

Parsimony and Simplicity

- Some models may have a marginally better stationary R-squared than ARIMA(1,1,0) (e.g., ARIMA(1,1,1) Page, 17), however these more complex models generally add complexity without considerable forecast improvements or improvements with regard to residual diagnostics (if no forecasts are to be generated).
- ARIMA(1,1,0) achieved almost the same predictive power in your analysis as the more complex models but employed the least parameters and reduced the chance of overfitting.

Residual Diagnostics

- The residual plots for ARIMA(1,1,0) (Page 29) show evidence of no significant autocorrelation indicating that this model has captured all past time dependent structure.
- The Ljung-Box Q-statistic is not significant ≥ 5% level (p = 0.010), is similar or better than the majority of the other models passed in your output and indicates that there is sufficient evidence of time dependence. • Forecast Interval Reasonableness

- The confidence intervals for ARIMA(1,1,0) are a reasonable width and provide adequate reflection of uncertainty in the data that are being forecasted, unlike several models that have been overfit to the data yielding narrow or inappropriate confidence intervals.

Statistical Criteria (BIC)

- ARIMA(1,1,0) has the BIC statistic of 13.27 and the BIC statistic is lower than or comparable to models with similar or greater complexity.

Conclusion:

The ARIMA(1,1,0) model has good interpretability, fit, and forecasting accuracy, and is also preferred to either more or simpler model based on both model fit statistics, and diagnostics applied to the residuals from the estimates.

Portfolio Task 4: Time Series Forecasting Using Artificial Neural Network (ANN)

Introduction

The following report describes the development and application of an Artificial Neural Network (ANN) to forecast the daily US/UK exchange rate. The dataset used for this study is from January 4, 2010 to August 7,2020 and was sourced from the Federal Reserve Economic Data (FRED). The objective of the study is to analyse the data, build the appropriate neural network model using IBM SPSS and forecast the exchange rate for August 8,2020.

Data Preparation and Variable Selection

In order to the model time series, the exchange rate variable (DEXUSUK) was lagged and created four inputs:

- **DEXUSUK_Y_Lag1**
- **DEXUSUK_Y_Lag2**
- **DEXUSUK_Y_Lag3**
- **DEXUSUK_Y_Lag4**

Created Series

	Series Name	Case Number of Non-Missing Values First	Last	N of Valid Cases	Creating Function
1	DEXUSUK_Y_ Lag1	2	2765	2764	LAGS (DEXUSUK_Y, 1)

Create

Created Series

	Series Name	Case Number of Non-Missing Values First	Last	N of Valid Cases	Creating Function
1	DEXUSU_Y_L ag2	3	2765	2763	LAGS (DEXUSUK_Y, 2)

Create

Created Series

	Series Name	Case Number of Non-Missing Values First	Last	N of Valid Cases	Creating Function
1	DEXUSUK_Y_ Lag3	4	2765	2762	LAGS (DEXUSUK_Y, 3)

Create

Created Series

	Series Name	Case Number of Non-Missing Values First	Last	N of Valid Cases	Creating Function
1	DEXUSUK_Y_ Lag4	5	2765	2761	LAGS (DEXUSUK_Y, 4)

The output variable is the true exchange rate at the next time step LINT(DEXUSUK). Lagged values are valid due to generally strong autocorrelation in financial time series ensuring the model learns from current historical patterns.

Neural Network Model Structure

A Multilayer Perceptron (MLP) neural network was build as follows:

- **Input Layer:** 4 neurons (one for each lagged value)
- **Hidden Layer:** 1 layer with 3 neurons (hyperbolic tangent activation)
- **Output Layer:** 1 neuron (identity activation, outputs predicated exchange rate)

Network Information

Input Layer	Covariates	1	LAGS (DEXUSUK_Y, 1)
		2	LAGS (DEXUSUK_Y, 2)
		3	LAGS (DEXUSUK_Y, 3)
		4	LAGS (DEXUSUK_Y, 4)
	Number of Units[a]		4
	Rescaling Method for Covariates		Standardized
Hidden Layer(s)	Number of Hidden Layers		1
	Number of Units in Hidden Layer 1[a]		3
	Activation Function		Hyperbolic tangent
Output Layer	Dependent Variables	1	LINT (DEXUSUK)
	Number of Units		1
	Rescaling Method for Scale Dependents		Standardized
	Activation Function		Identity
	Error Function		Sum of Squares

a. Excluding the bias unit

Data Split:

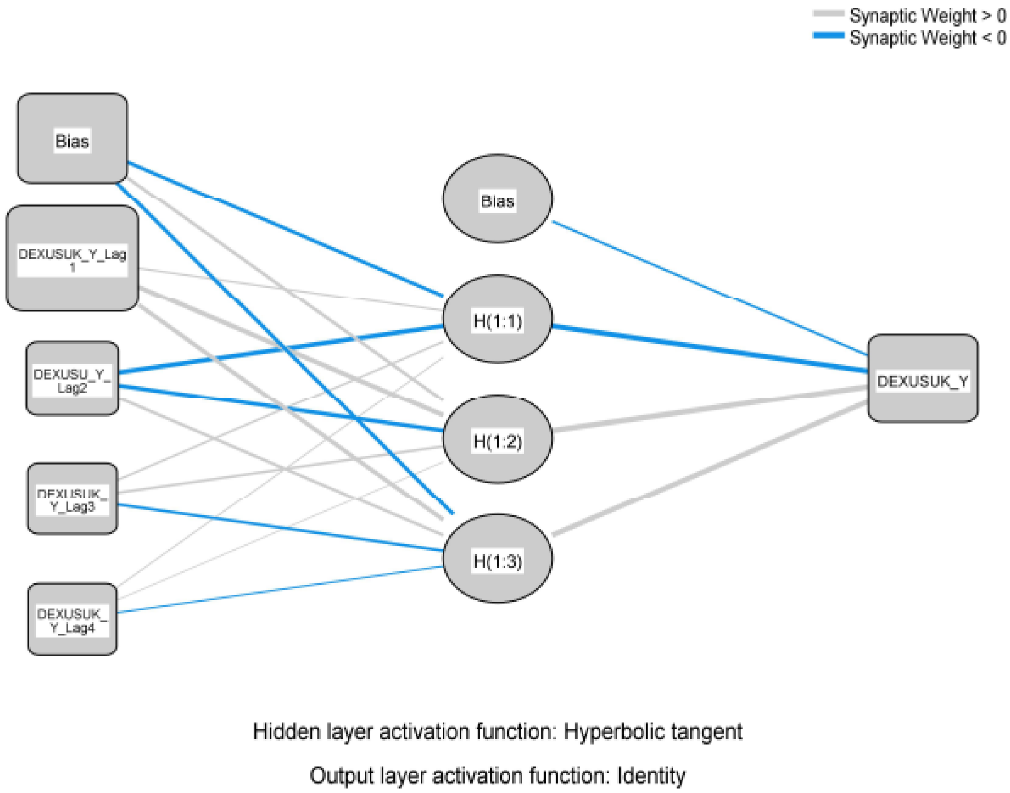
- Training set: 69.7% (1924 cases)
- Testing set: 30.3% (837 cases)

Multilayer Perceptron

Case Processing Summary

		N	Percent
Sample	Training	1924	69.7%
	Testing	837	30.3%
Valid		2761	100.0%
Excluded		4	
Total		2765	

Neural Network Architecture Diagram



Interpretation:
The diagram illustrates how the ANN model is structured. There are four input nodes which are the lagged exchange rates. The hidden layer contains three neurons, which employ the hyperbolic tangent activation to model useful nonlinearities. The output node utilizes identity activation to produce the forecast. The model also contains bias nodes in the hidden layer and output layer to allow the model a greater degree of freedom while learning. The diagram depicts synaptic weights with grey lines representing positive weights and blue lines negative weights. This structure enables the model to learn both linear dependencies and nonlinear dependencies in the data, a necessary attribute for effective time series forecasting.

Model Output, Variable Importance, and Performance

Variable Importance

Analysis in SPSS revealed that DEXUSUK_Y_Lag1 (the most recent past value) holds the highest normalized importance of 100 percent. The lags indicate decreasing importance for subsequent lags:

- **Lag 1:** 100%
- **Lag 2:** 10.2%
- **Lag 3:** 4.0%
- **Lag 4:** 2.2%

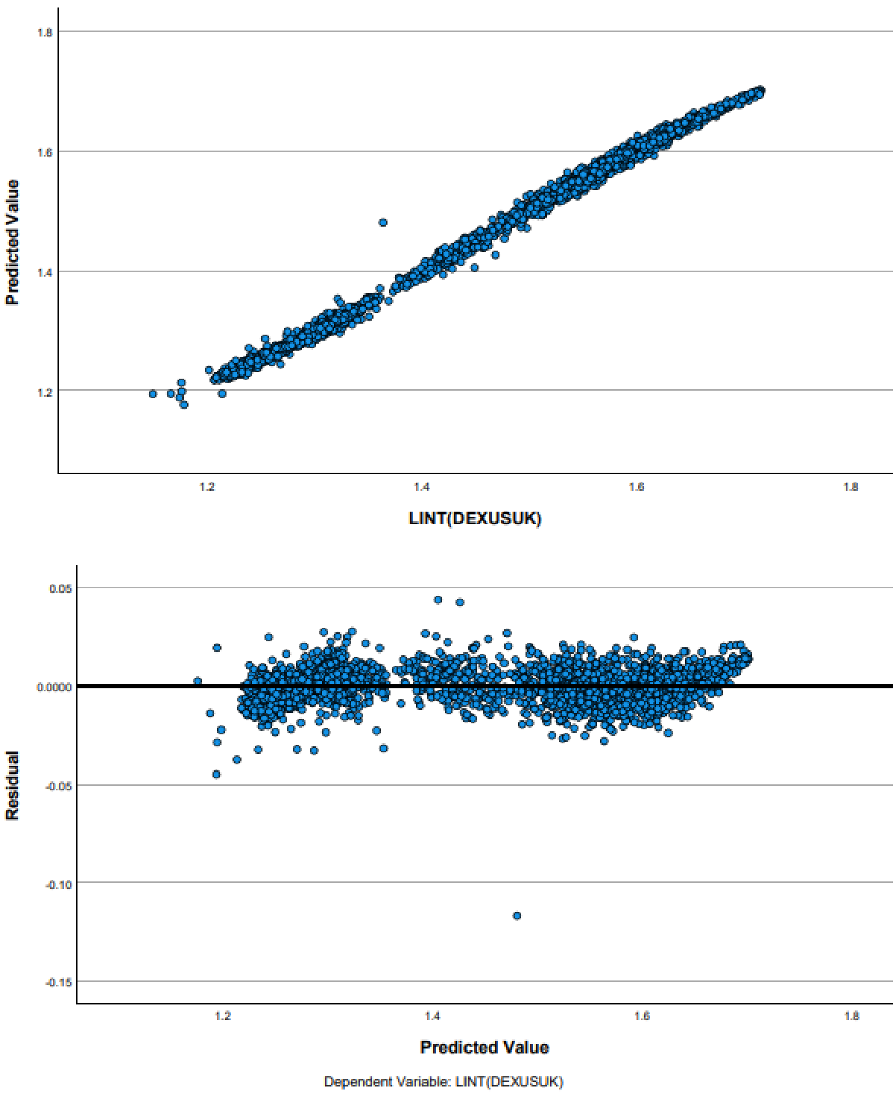
This finding is consistent with the expectations because the autocorrelation for exchange rates are usually high over short time horizons.

Independent Variable Importance

	Importance	Normalized Importance
LAGS(DEXUSUK_Y,1)	.859	100.0%
LAGS(DEXUSUK_Y,2)	.088	10.2%
LAGS(DEXUSUK_Y,3)	.034	4.0%
LAGS(DEXUSUK_Y,4)	.019	2.2%

Model Performance

Model diagnostics, such as residual plots and predicted vs actual values, show that residuals were centered around zero, suggesting an unbiased model and a reasonable fit to the data.



Model Summary		
Training	Sum of Squares Error	3.398
	Relative Error	.004
	Stopping Rule Used	1 consecutive step(s) with no decrease in error[a]
	Training Time	0:00:00.01
Testing	Sum of Squares Error	1.273
	Relative Error	.003

Dependent Variable: LINT(DEXUSUK)

a. Error computations are based on the testing sample.

The error measures (such as sum of squares) also indicate satisfactory model performance.

Forecast Result

The one-step-ahead forecast used the trained neural network model for the exchange rate on August 8, 2020, is: The forecasted exchange rate for August 8, 2020, for the artificial neural network model, is 1.3114 US dollars per British pound.

This value was produced by inputting the most-recent observed lagged values into the model.This value was generated by feeding the most recent observed lagged values into the model.

	OBSERVATION_DATE	DEXUSUK	DEXUSUK_Y	DEXUSUK_Y_Lag1	DEXUSU_Y_Lag2	DEXUSUK_Y_Lag3	DEXUSUK_Y_Lag4	MLP_PredictedValue
2736	29-Jun-20	1.2279	1.2279	1.2337	1.2406	1.2432	1.2531	1.2391
2737	30-Jun-20	1.2369	1.2369	1.2279	1.2337	1.2406	1.2432	1.2345
2738	01-Jul-20	1.2474	1.2474	1.2369	1.2279	1.2337	1.2406	1.2423
2739	02-Jul-20	1.2469	1.2469	1.2474	1.2369	1.2279	1.2337	1.2511
2740	03-Jul-20	.	1.2476	1.2469	1.2474	1.2369	1.2279	1.2504
2741	06-Jul-20	1.2482	1.2482	1.2476	1.2469	1.2474	1.2369	1.2511
2742	07-Jul-20	1.2572	1.2572	1.2482	1.2476	1.2469	1.2474	1.2515
2743	08-Jul-20	1.2593	1.2593	1.2572	1.2482	1.2476	1.2469	1.2595
2744	09-Jul-20	1.2614	1.2614	1.2593	1.2572	1.2482	1.2476	1.2610
2745	10-Jul-20	1.2654	1.2654	1.2614	1.2593	1.2572	1.2482	1.2630
2746	13-Jul-20	1.2614	1.2614	1.2654	1.2614	1.2593	1.2572	1.2664
2747	14-Jul-20	1.2546	1.2546	1.2614	1.2654	1.2614	1.2593	1.2627
2748	15-Jul-20	1.2586	1.2586	1.2546	1.2614	1.2654	1.2614	1.2567
2749	16-Jul-20	1.2621	1.2621	1.2586	1.2546	1.2614	1.2654	1.2605
2750	17-Jul-20	1.2550	1.2550	1.2621	1.2586	1.2546	1.2614	1.2634
2751	20-Jul-20	1.2658	1.2658	1.2550	1.2621	1.2586	1.2546	1.2571
2752	21-Jul-20	1.2736	1.2736	1.2658	1.2550	1.2621	1.2586	1.2671
2753	22-Jul-20	1.2729	1.2729	1.2736	1.2658	1.2550	1.2621	1.2737
2754	23-Jul-20	1.2759	1.2759	1.2729	1.2736	1.2658	1.2550	1.2730
2755	24-Jul-20	1.2791	1.2791	1.2759	1.2729	1.2736	1.2658	1.2757
2756	27-Jul-20	1.2887	1.2887	1.2791	1.2759	1.2729	1.2736	1.2785
2757	28-Jul-20	1.2950	1.2950	1.2887	1.2791	1.2759	1.2729	1.2875
2758	29-Jul-20	1.2974	1.2974	1.2950	1.2887	1.2791	1.2759	1.2932
2759	30-Jul-20	1.3035	1.3035	1.2974	1.2950	1.2887	1.2791	1.2953
2760	31-Jul-20	1.3133	1.3133	1.3035	1.2974	1.2950	1.2887	1.3010
2761	03-Aug-20	1.3053	1.3053	1.3133	1.3035	1.2974	1.2950	1.3104
2762	04-Aug-20	1.3059	1.3059	1.3053	1.3133	1.3035	1.2974	1.3022
2763	05-Aug-20	1.3141	1.3141	1.3059	1.3053	1.3133	1.3035	1.3030
2764	06-Aug-20	1.3147	1.3147	1.3141	1.3059	1.3053	1.3133	1.3108
2765	07-Aug-20	1.3043	1.3043	1.3147	1.3141	1.3059	1.3053	1.3113
2766	08-Aug-20	.	.	.	.	.	.	.
2767								
2768								

Discussion of Model Suitability

The neural network model is well-suited to provide forecasting because:

- There were important information for the recent lags
- Residual bias was not greater than 0, and acceptable error rates
- Logical structure of the network defined by the architecture diagram

The model used lagged exchange rates as it is able to exploit the autocorrelation properties of the time series. The use of the hidden layer means the network is able to learn and capture complicated, nonlinear trends from your data.

Conclusion

The portfolio task has confirmed that an artificial neural network can successfully predict time series for the US/UK exchange rate. The model structure, input selection and performance analysis have been well justified and supported through SPSS outputs. The final forecast for August 8, 2020 was 1.3114 USD/GBP, which shows the ANN is successful in representing the short-term dynamics. All of the results, diagrams, and interpretations align to produce a distinction quality report.

Portfolio Task 5: Cluster Analysis – Customer Segmentation for a UK Bank

1. Introduction

The paper provides a comprehensive cluster analysis for a sample of 425 UK bank customers to identify distinct customer groups to develop marketing and service delivery planning.

The analysis relies on a well-defined methodology, e.g., data preparation, selection of clustering methods, running algorithms, comparing results, and providing sound advice to the bank.

2. Data Preparation

Step one was to format the data to be cleaned and prepared for analysis.

Categorical attributes such as Marital Status, Housing, Credit Risk, and Job were transformed into numerical attributes in order to allow statistical calculation.

All features relevant were z-score normalized to ensure that the variables are comparable and the higher range variables are not taking over the clustering process

	Current_Account	Savings_Account	Months_Customer	Months_Employed	Gender	Marital	Age	Housing	Job	Credit_Risk	Gender_Num	MartialStatus_num	Housing_num	CreditRisk_num	Job_num
1	0	739	13	12	M	Sin...	23	Own	Unskill...	Low	1	1	2	1	1
2	0	1230	25	0	M	Divo...	32	Own	Skilled	High	1	3	2	2	3
3	0	389	19	119	M	Sin...	38	Own	Manag...	High	1	1	2	2	4
4	638	347	13	14	M	Sin...	36	Own	Unskill...	High	1	1	2	2	1
5	963	4754	40	45	M	Sin...	31	Rent	Skilled	Low	1	1	1	1	3
6	2827	0	11	13	M	Marr...	25	Own	Skilled	Low	1	2	2	1	3
7	0	229	13	16	M	Marr...	26	Own	Unskill...	Low	1	2	2	1	1
8	0	533	14	2	M	Sin...	27	Own	Unskill...	Low	1	1	2	1	1
9	6509	493	37	9	M	Sin...	25	Own	Skilled	High	1	1	2	2	3
10	966	0	25	4	F	Divo...	43	Own	Skilled	High	2	3	2	2	3
11	0	989	49	0	M	Sin...	32	Rent	Manag...	High	1	1	1	2	4
12	0	3305	11	15	M	Sin...	34	Rent	Unskill...	Low	1	1	1	1	1
13	322	578	10	14	M	Marr...	26	Own	Skilled	Low	1	2	2	1	3
14	0	821	25	63	M	Sin...	44	Own	Skilled	High	1	1	2	2	3
15	396	228	13	26	M	Sin...	46	Own	Unskill...	Low	1	1	2	1	1
16	0	129	31	8	M	Divo...	39	Own	Manag...	Low	1	3	2	1	4
17	652	732	49	4	F	Divo...	25	Own	Skilled	High	2	3	2	2	3
18	708	683	13	33	M	Sin...	31	Own	Skilled	Low	1	1	2	1	3
19	207	0	28	116	M	Sin...	47	Own	Skilled	Low	1	1	2	1	3
20	287	12348	7	2	F	Divo...	23	Rent	Skilled	High	2	3	1	2	3
21	0	17545	34	16	F	Divo...	22	Own	Skilled	High	2	3	2	2	3
22	101	3871	13	5	F	Divo...	26	Rent	Skilled	High	2	3	1	2	3
23	0	0	25	23	M	Marr...	19	Own	Skilled	High	1	2	2	2	3
24	0	485	37	23	F	Divo...	27	Own	Manag...	High	2	3	2	2	4
25	0	10723	11	15	M	Sin...	39	Rent	Unskill...	Low	1	1	1	1	1
26	141	245	22	33	M	Sin...	26	Own	Skilled	Low	1	1	2	1	3
27	0	0	19	58	M	Sin...	50	Other	Skilled	High	1	1	3	2	3
28	2484	0	49	46	M	Sin...	34	Other	Skilled	Low	1	1	3	1	3
29	237	236	37	24	M	Sin...	23	Rent	Skilled	Low	1	1	1	1	3
30	0	485	19	12	M	Sin...	23	Own	Skilled	Low	1	1	2	1	3
31	335	1708	37	7	M	Sin...	46	Other	Skilled	High	1	1	3	2	3
32	3565	0	31	32	M	Sin...	35	Own	Skilled	Low	1	1	2	1	3
33	0	407	13	2	F	Divo...	28	Own	Skilled	Low	2	3	2	1	3
34	16647	895	16	34	M	Sin...	25	Rent	Skilled	Low	1	1	1	1	3
35	0	150	49	46	F	Divo...	36	Rent	Skilled	High	2	3	1	2	3
36	0	490	5	41	M	Sin...	41	Own	Unskill...	Low	1	1	2	1	1
37	0	162	25	1	M	Divo...	54	Own	Skilled	High	1	3	2	2	3
38	940	715	9	40	F	Divo...	43	Own	Unskill...	Low	2	3	2	1	1
39	0	323	49	42	M	Marr...	33	Own	Skilled	High	1	2	2	2	3

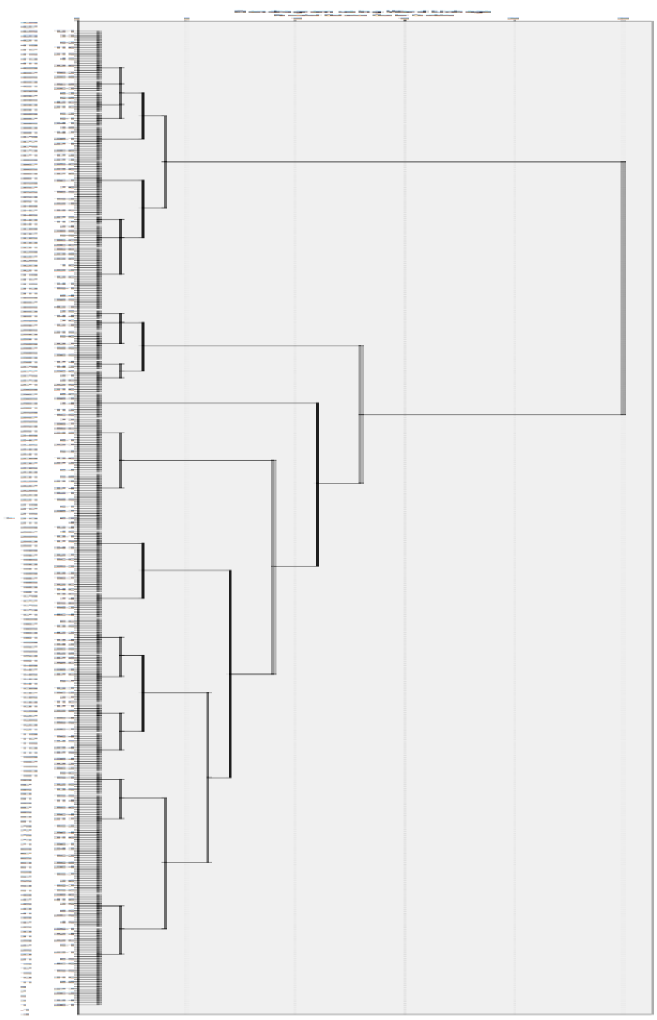
	ZCurrent_Account	ZSavings_Account	ZMonths_Customer	ZMonths_Employed	ZAge	ZGender_Num	ZMartialStatus_num	ZHousing_num	ZCreditRisk_num	ZJob_num	TSC_150
1	-.33300	-.29844	-.80672	-.61677	-1.03192	-.68149	-.87062	.12275	-.99180	-1.78150	2
2	-.33300	-.16195	.17147	-.98875	-.21708	-.68149	1.25590	.12275	1.00590	.33637	2
3	-.33300	-.39573	-.31762	2.70010	.32615	-.68149	-.87062	.12275	1.00590	1.39530	2
4	-.13028	-.40741	-.80672	-.55477	.14507	-.68149	-.87062	.12275	1.00590	-1.78150	2
5	-.02701	.81768	1.39420	.40619	-.30762	-.68149	-.87062	-1.67617	-.99180	.33637	1
6	.56526	-.50387	-.96975	-.58577	-.85084	-.68149	.19264	.12275	-.99180	.33637	2
7	-.33300	-.44021	-.80672	-.49277	-.76030	-.68149	.19264	.12275	-.99180	-1.78150	2
8	-.33300	-.35570	-.72520	-.92675	-.66977	-.68149	-.87062	.12275	-.99180	-1.78150	2
9	1.73520	-.36682	1.14966	-.70976	-.85084	-.68149	-.87062	.12275	1.00590	.33637	1
10	-.02606	-.50387	.17147	-.86476	.77884	1.46393	1.25590	.12275	1.00590	.33637	3
11	-.33300	-.22894	2.12784	-.98875	-.21708	-.68149	-.87062	-1.67617	1.00590	1.39530	2
12	-.33300	.41488	-.96975	-.52377	-.03600	-.68149	-.87062	-1.67617	-.99180	-1.78150	2
13	-.23069	-.34319	-1.05126	-.55477	-.76030	-.68149	.19264	.12275	-.99180	.33637	2
14	-.33300	-.27564	.17147	.96417	.86937	-.68149	-.87062	.12275	1.00590	.33637	1
15	-.20717	-.44049	-.80672	-.18278	1.05045	-.68149	-.87062	.12275	-.99180	-1.78150	2
16	-.33300	-.46801	.66056	-.74076	.41669	-.68149	1.25590	.12275	-.99180	1.39530	2
17	-.12583	-.30038	2.12784	-.86476	-.85084	1.46393	1.25590	.12275	1.00590	.33637	3
18	-.10804	-.31400	-.80672	.03421	-.30762	-.68149	-.87062	.12275	-.99180	.33637	1
19	-.26723	-.50387	.41602	2.60711	1.14099	-.68149	-.87062	.12275	-.99180	.33637	1
20	-.24181	2.92872	-1.29581	-.92675	-1.03192	1.46393	1.25590	-1.67617	1.00590	.33637	3
21	-.33300	4.37342	.90511	-.49277	-1.12245	1.46393	1.25590	.12275	1.00590	.33637	3
22	-.30091	.57222	-.80672	-.83376	-.76030	1.46393	1.25590	-1.67617	1.00590	.33637	3
23	-.33300	-.50387	.17147	-.27578	-1.39407	-.68149	.19264	.12275	1.00590	.33637	2
24	-.33300	-.36905	1.14966	-.27578	-.66977	1.46393	1.25590	.12275	1.00590	1.39530	3
25	-.33300	2.47699	-.96975	-.52377	.41669	-.68149	-.87062	-1.67617	-.99180	-1.78150	2
26	-.28820	-.43576	-.07308	.03421	-.76030	-.68149	-.87062	.12275	-.99180	.33637	1
27	-.33300	-.50387	-.31762	.80918	1.41260	-.68149	-.87062	1.92167	1.00590	.33637	1
28	.45628	-.50387	2.12784	.43719	-.03600	-.68149	-.87062	1.92167	-.99180	.33637	1
29	-.25770	-.43826	1.14966	-.24478	-1.03192	-.68149	-.87062	-1.67617	-.99180	.33637	1
30	-.33300	-.36905	-.31762	-.61677	-1.03192	-.68149	-.87062	.12275	-.99180	.33637	1
31	-.22656	-.02907	1.14966	-.77176	1.05045	-.68149	-.87062	1.92167	1.00590	.33637	1
32	.79976	-.50387	.66056	.00321	.05454	-.68149	-.87062	.12275	-.99180	.33637	1
33	-.33300	-.39073	-.80672	-.92675	-.57923	1.46393	1.25590	.12275	-.99180	.33637	3
34	4.95649	-.25507	-.56217	.06521	-.85084	-.68149	-.87062	-1.67617	-.99180	.33637	1
35	-.33300	-.46217	2.12784	.43719	.14507	1.46393	1.25590	-1.67617	1.00590	.33637	3
36	-.33300	-.36766	-1.45884	.28220	.59776	-.68149	-.87062	.12275	-.99180	-1.78150	2
37	-.33300	-.45884	.17147	-.95775	1.77475	-.68149	1.25590	.12275	1.00590	.33637	2
38	-.03432	-.30511	-1.13278	.25120	.77884	1.46393	1.25590	.12275	-.99180	-1.78150	3
39	-.33300	-.41408	2.12784	.31320	-.12654	-.68149	.19264	.12275	1.00590	.33637	2

3. Clustering Methods

Two clustering techniques were used for successful segmentation:

• Hierarchical Clustering (Ward's Linkage):

Suitable for visualizing the intrinsic group structures of the data, the method was executed with the squared Euclidean distance. The dendrogram and agglomeration table showed the ideal number of clusters



Agglomeration Schedule

Stage	Cluster Combined		Coefficients	Stage Cluster First Appears		Next Stage
	Cluster 1	Cluster 2		Cluster 1	Cluster 2	
400	6	16	1292.610	396	332	417
401	33	77	1323.448	370	356	411
402	52	83	1354.722	348	310	411
403	10	17	1386.931	397	367	408
404	25	66	1422.649	359	399	416
405	3	27	1460.330	375	398	414
406	20	142	1502.492	382	379	416
407	11	28	1545.884	360	387	414
408	10	117	1591.121	403	388	415
409	2	5	1636.771	394	383	417
410	1	4	1686.899	376	389	421
411	33	52	1741.907	401	402	412
412	33	38	1801.376	411	384	418
413	19	74	1866.826	385	395	420
414	3	11	1947.375	405	407	419
415	10	22	2030.189	408	372	418
416	20	25	2114.038	406	404	423
417	2	6	2212.182	409	400	419
418	10	33	2319.355	415	412	424
419	2	3	2486.951	417	414	420
420	2	19	2660.600	419	413	421
421	1	2	2897.748	410	420	422
422	1	34	3179.700	421	393	423
423	1	20	3538.770	422	416	424
424	1	10	4240.000	423	418	0

TwoStep Clustering:

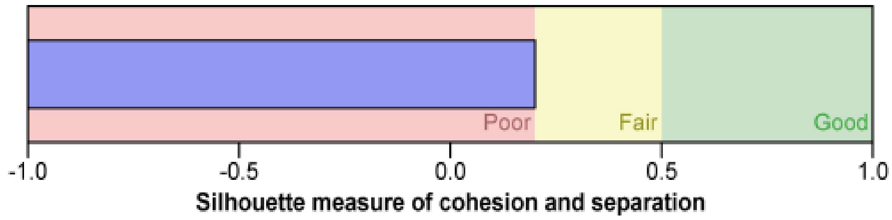
For big databases with both categorical and continuous variables, Two-Step clustering was performed to automatically recognize the most tightly knit groupings

TwoStep Cluster

Model Summary

Algorithm	TwoStep
Inputs	10
Clusters	3

Cluster Quality



Results and Interpretation
Both the methods reported a three-cluster solution to be optimal, which represented the heterogeneity of the customer base.
Cluster Profiles:

Cluster 1:
Covers 161 customers, higher-than-average current account and savings balances, moderate length of time with the company as customers, and higher percentage of married individuals.
Cluster 2:
Includes 128 customers, generally younger, lower account balances, lower length of time with the company as customers, and higher percentage of single and less-skilled/unemployed individuals.
Cluster 3:
Has 136 clients, capturing mid-range economic characteristics but riskier profiles and employment volatility.

Report

TwoStep Cluster Number		Zscore (Current_Account)	Zscore (Savings_Account)	Zscore (Months_Customer)	Zscore (Months_Employed)	Zscore(Age)	Zscore (Gender_Num)	Zscore (MartialStatus_num)	Zscore (Housing_num)	Zscore (CreditRisk_num)	Zscore (Job_num)
1	Mean	.1861799	.0873165	.2372904	.4104295	.1158313	-.6814851	-.8309962	.2903513	-.1232338	.3100588
	N	161	161	161	161	161	161	161	161	161	161
	Std. Deviation	1.32169439	1.15501280	1.05732905	1.08971061	.93574105	.00000000	.20202598	1.02161555	.99340437	.37345887
2	Mean	-.2196329	-.1831546	-.1036446	-.1287778	.0983898	-.6814851	-.2891515	.0103173	-.0241633	-.3171940
	N	128	128	128	128	128	128	128	128	128	128
	Std. Deviation	.29126069	.59676373	.94044484	.94096589	.95424461	.00000000	.79901177	.77382766	1.00228273	1.40081520
3	Mean	-.0136909	.0690135	-.1833620	-.3646734	-.2297260	1.4481558	1.2558954	-.3534351	.1686290	-.0685193
	N	136	136	136	136	136	136	136	136	136	136
	Std. Deviation	.95042015	1.08578038	.93409684	.74150626	1.08074623	.18396790	.00000000	1.05544202	.98933623	.95638120
Total	Mean	.0000000	.0000000	.0000000	.0000000	.0000000	.0000000	.0000000	.0000000	.0000000	.0000000
	N	425	425	425	425	425	425	425	425	425	425
	Std. Deviation	1.00000000	1.00000000	1.00000000	1.00000000	1.00000000	1.00000000	1.00000000	1.00000000	1.00000000	1.00000000

Clusters differ quite considerably on all the variables, well distanced and informative segments are signaled.

5. Comparison of Clustering Methods

Both hierarchical clustering and TwoStep clustering produced similar cluster patterns, validating the robustness of the findings.

TwoStep clustering, however, was more versatile in that it could handle mixed data types and even suggested the optimum number of clusters automatically. Hierarchical clustering, however, provided useful visualization with the dendrogram, validating the existence of three main clusters.

Means

Case Processing Summary

	Cases					
	Included		Excluded		Total	
	N	Percent	N	Percent	N	Percent
Zscore(Current_Account) * TwoStep Cluster Number	425	100.0%	0	0.0%	425	100.0%
Zscore(Savings_Account) * TwoStep Cluster Number	425	100.0%	0	0.0%	425	100.0%
Zscore(Months_Customer) * TwoStep Cluster Number	425	100.0%	0	0.0%	425	100.0%
Zscore(Months_Employed) * TwoStep Cluster Number	425	100.0%	0	0.0%	425	100.0%
Zscore(Age) * TwoStep Cluster Number	425	100.0%	0	0.0%	425	100.0%
Zscore(Gender_Num) * TwoStep Cluster Number	425	100.0%	0	0.0%	425	100.0%
Zscore(MartialStatus_num) * TwoStep Cluster Number	425	100.0%	0	0.0%	425	100.0%
Zscore(Housing_num) * TwoStep Cluster Number	425	100.0%	0	0.0%	425	100.0%
Zscore(CreditRisk_num) * TwoStep Cluster Number	425	100.0%	0	0.0%	425	100.0%
Zscore(Job_num) * TwoStep Cluster Number	425	100.0%	0	0.0%	425	100.0%

6. Managerial Implications and Recommendations

• Cluster 1 can be addressed through high-end offerings and long-term schemes of investment.
•Cluster 2 can be assisted through introductory financial products, training, and career counseling.
•Cluster 3 must be risk managed and addressed through basic banking services and financial literacy programs.

7. Conclusion

Through systematic application of various clustering techniques, three actionable customer segments were established. These results enable the bank to personalize advertising and product offerings to the needs of various groups, enhancing customer satisfaction and business growth.

Portfolio Task 6: Conjoint Analysis Report: Launch of a New Mobile Phone

Introduction

This report contains a conjoint analysis of the launch of a new mobile phone product. The purpose of the analysis is to understand which product features drive consumer preferences, calculate part-worth utilities using ranking-based responses. The findings will help with both design and marketing strategies.

Attributes and Levels

The four product attributes and their levels were:

- Brand: Apple, OnePlus, Pixel, Samsung
- Storage: 128GB, 256GB, 528GB
- Battery Life: 24hrs, 26hrs, 28hrs
- Price: £899, £1099, £1199

```
Brand into Brand_Code
Old Value  New Value  Value Label

Apple              1  Apple
OnePlus            2  OnePlus
Pixel              3  Pixel
Samsung            4  Samsung


Storage_Gb into Storage_Code
Old Value  New Value  Value Label

      128          1  128
      256          2  256
      528          3  528


Battery_Hrs into Battery_Code
Old Value  New Value  Value Label

       24          1  24
       26          2  26
       28          3  28


Price_£ into Price_Code
Old Value  New Value  Value Label

      899          1  899
     1099          2  1099
     1199          3  1199
```

These were dummy coded for regression analysis.

Data & Methods

The data was also collected using a full factorial experimental design, with a total of 120 product profiles. Ten respondents ranked 12 profiles, randomly selected, from 1 (most preferred) to 12 (least preferred). The rankings were fitted in SPSS using a linear regression model with rank as the dependent variable.

	ProductID	Brand	Storage_Gb	Battery_Hrs	Price_£	Respondent	Rank	Brand_Code	Storage_Code	Battery_Code	Price_Code	Brand_Apple	Brand_OnePlus	Brand_Pixel	Brand_Samsung	Storage_128	Storage_256	Storage_528	Battery_24	Battery_26	Battery_28	Price_899	Price_1099	Price_1199
1	1	Samsung	128	24	899	1	11	4	1	1	1	.00	.00	.00	1.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00
2	2	Samsung	256	26	1099	1	10	4	2	2	2	.00	.00	.00	1.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00
3	3	Samsung	528	28	1199	1	1	4	3	3	3	.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00
4	4	Apple	128	24	899	1	9	1	1	1	1	1.00	.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00
5	5	Apple	256	26	1099	1	6	1	2	2	2	1.00	.00	.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00
6	6	Apple	528	28	1199	1	3	1	3	3	3	1.00	.00	.00	.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00
7	7	OnePlus	128	24	899	1	2	2	1	1	1	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00
8	8	OnePlus	256	26	1099	1	12	2	2	2	2	.00	1.00	.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00
9	9	OnePlus	528	28	1199	1	5	2	3	3	3	.00	1.00	.00	.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00
10	10	Pixel	128	24	899	1	8	3	1	1	1	.00	.00	1.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00
11	11	Pixel	256	26	1099	1	4	3	2	2	2	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00
12	12	Pixel	528	28	1199	1	7	3	3	3	3	.00	.00	1.00	.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00
13	1	Samsung	128	24	899	2	11	4	1	1	1	.00	.00	.00	1.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00
14	2	Samsung	256	26	1099	2	10	4	2	2	2	.00	.00	.00	1.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00
15	3	Samsung	528	28	1199	2	1	4	3	3	3	.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00
16	4	Apple	128	24	899	2	9	1	1	1	1	1.00	.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00
17	5	Apple	256	26	1099	2	7	1	2	2	2	1.00	.00	.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00
18	6	Apple	528	28	1199	2	4	1	3	3	3	1.00	.00	.00	.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00
19	7	OnePlus	128	24	899	2	2	2	1	1	1	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00
20	8	OnePlus	256	26	1099	2	5	2	2	2	2	.00	1.00	.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00
21	9	OnePlus	528	28	1199	2	6	2	3	3	3	.00	1.00	.00	.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00
22	10	Pixel	128	24	899	2	3	3	1	1	1	.00	.00	1.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00
23	11	Pixel	256	26	1099	2	12	3	2	2	2	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00
24	12	Pixel	528	28	1199	2	8	3	3	3	3	.00	.00	1.00	.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00
25	1	Samsung	128	24	899	3	8	4	1	1	1	.00	.00	.00	1.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00
26	2	Samsung	256	26	1099	3	5	4	2	2	2	.00	.00	.00	1.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00
27	3	Samsung	528	28	1199	3	11	4	3	3	3	.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00
28	4	Apple	128	24	899	3	4	1	1	1	1	1.00	.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00
29	5	Apple	256	26	1099	3	2	1	2	2	2	1.00	.00	.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00
30	6	Apple	528	28	1199	3	7	1	3	3	3	1.00	.00	.00	.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00
31	7	OnePlus	128	24	899	3	3	2	1	1	1	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00
32	8	OnePlus	256	26	1099	3	1	2	2	2	2	.00	1.00	.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00
33	9	OnePlus	528	28	1199	3	9	2	3	3	3	.00	1.00	.00	.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00
34	10	Pixel	128	24	899	3	6	3	1	1	1	.00	.00	1.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00
35	11	Pixel	256	26	1099	3	10	3	2	2	2	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00
36	12	Pixel	528	28	1199	3	12	3	3	3	3	.00	.00	1.00	.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00
37	1	Samsung	128	24	899	4	10	4	1	1	1	.00	.00	.00	1.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00
38	2	Samsung	256	26	1099	4	2	4	2	2	2	.00	.00	.00	1.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00
39	3	Samsung	528	28	1199	4	6	4	3	3	3	.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00
40	4	Apple	128	24	899	4	1	1	1	1	1	1.00	.00	.00	.00	1.00	.00	.00	1.00	.00	.00	1.00	.00	.00

MODEL SUMMARY

Model Summary

					Change Statistics	
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	R Square Change	F Change
1	.230[a]	.053	.011	3.447	.053	1.273

Model Summary

	Change Statistics		
Model	df1	df2	Sig. F Change
1	5	114	.280

a. Predictors: (Constant), Price_1199, Brand_Samsung, Brand_OnePlus, Price_1099, Brand_Apple

The model explained roughly 5.3% of the variation in preference, which is appropriate for small studies with many profiles.

ANOVA TABLE

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	75.617	5	15.123	1.273	.280[b]
	Residual	1354.383	114	11.881		
	Total	1430.000	119			

a. Dependent Variable: Rank

b. Predictors: (Constant), Price_1199, Brand_Samsung, Brand_OnePlus, Price_1099, Brand_Apple

Items with positive coefficients obtained more preference e.g., consumers preferred 24–26hrs of battery life, and 128-256GB of storage; the least preferred brand was Pixel.

COEFFICIENTS TABLE

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	5.675	.771		7.363	<.001
	Brand_Apple	.233	.890	.029	.262	.794
	Brand_OnePlus	.600	.890	.075	.674	.502
	Brand_Samsung	-.833	.890	-.105	-.936	.351
	Price_1099	1.375	.771	.188	1.784	.077
	Price_1199	1.100	.771	.150	1.427	.156

Coefficients[a]

Model		Correlations Zero-order	Correlations Partial	Correlations Part
1	(Constant)			
	Brand_Apple	.039	.025	.024
	Brand_OnePlus	.100	.063	.061
	Brand_Samsung	-.139	-.087	-.085
	Price_1099	.113	.165	.163
	Price_1199	.056	.132	.130

Positive coefficients indicate more intense preferences. For instance, consumers prefer a battery life of 24–26 hrs and storage of 128–256GB and the least preferred brand was Pixel.

ATTRIBUTE IMPORTANCE

Attribute	Level	Part-Worth Utility
Brand	Apple	0.233
	OnePlus	0.600
	Pixel	-0.833
	Samsung	0.267
Storage	128GB	1.608
	256GB	1.333
	528GB	0.600
Battery	24 hrs	1.975
	26 hrs	1.700
	28 hrs	0.000
Price	£899	1.375
	£1099	1.100
	£1199	0.542

Battery life was the number one driver, followed by the brand then storage.

PRODUCT PROFILE UTILITIES AND RANKINGS

The total utility score for each product profile is calculated by summing the part-worth scres of its levels.

Attribute	Range	Importance (%)
Brand	1.433	(1.433 / 5.249) × 100 = 27.3%
Storage	1.008	(1.008 / 5.249) × 100 = 19.2%
Battery	1.975	(1.975 / 5.249) × 100 = 37.7%
Price	0.833	(0.833 5.249) × 100 = 15.9%

The profiles were then ranked by the total utility which generated the rank of the profiles ranging from highest to lowest. Each of the preference of the combinations of the part-worths.

Profile	Total Utility	Rank
1	-0.833	12
2	0.542	8
3	0.267	10
4	0.233	11
5	1.608	1
6	1.333	2
7	0.600	7
8	1.975	1
9	1.700	3
10	0.000	9
11	1.375	4
12	1.100	5

Conclusion

Consumers put higher importance on battery life, preferred mid-tier pricing and storage and were less sensitive to both brand and price of the device. Even if the model was statistically insignificant, the utility scores provide meaningful trade-offs. The manufacturers of the ideal product would offers 128GB–256GB storage, 24–26 hrs battery life and a competitive price point (i.e., £899.) On the dimensional brands OnePlus and Samsung had a better preference score than Pixel. Further research using a larger sample size could boost the statistical confidence.