

Portfolio Project

Presenter:
Nawaz Quraishi

Portfolio Project

- ❏ Motivation
- ❏ Preview of the Data & Exploration
- ❏ Feature Engineering
- ❏ Modeling
- ❏ Next Steps
- ❏ Conclusion

Portfolio Project

Motivation: Perform Customer Segmentation and employ a Deep Learning Model on a Retail Data Set

Dataset:

- UCI Online II Retail Data Set
- 525,461 rows and 8 columns
- Transactions from 2009 to 2010

Portfolio Project

Preview of the dataset

	Invoice	StockCode	Description	Quantity	InvoiceDate	Price	Customer ID	Country
0	489434	85048	15CM CHRISTMAS GLASS BALL 20 LIGHTS	12	2009-12-01 07:45:00	6.95	13085.0	United Kingdom
1	489434	79323P	PINK CHERRY LIGHTS	12	2009-12-01 07:45:00	6.75	13085.0	United Kingdom
2	489434	79323W	WHITE CHERRY LIGHTS	12	2009-12-01 07:45:00	6.75	13085.0	United Kingdom
3	489434	22041	RECORD FRAME 7" SINGLE SIZE	48	2009-12-01 07:45:00	2.10	13085.0	United Kingdom
4	489434	21232	STRAWBERRY CERAMIC TRINKET BOX	24	2009-12-01 07:45:00	1.25	13085.0	United Kingdom

Portfolio Project

Messy data with number of missing values, negative values for quantities and price

	Quantity	Price	Customer ID
count	525461.000000	525461.000000	417534.000000
mean	10.337667	4.688834	15360.645478
std	107.424110	146.126914	1680.811316
min	-9600.000000	-53594.360000	12346.000000
25%	1.000000	1.250000	13983.000000
50%	3.000000	2.100000	15311.000000
75%	10.000000	4.210000	16799.000000
max	19152.000000	25111.090000	18287.000000

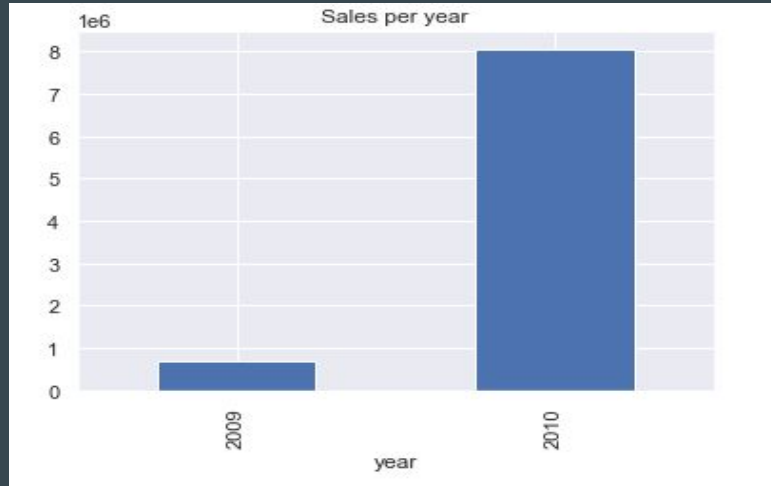
Portfolio Project

Exploratory Data Analysis - Heat Map



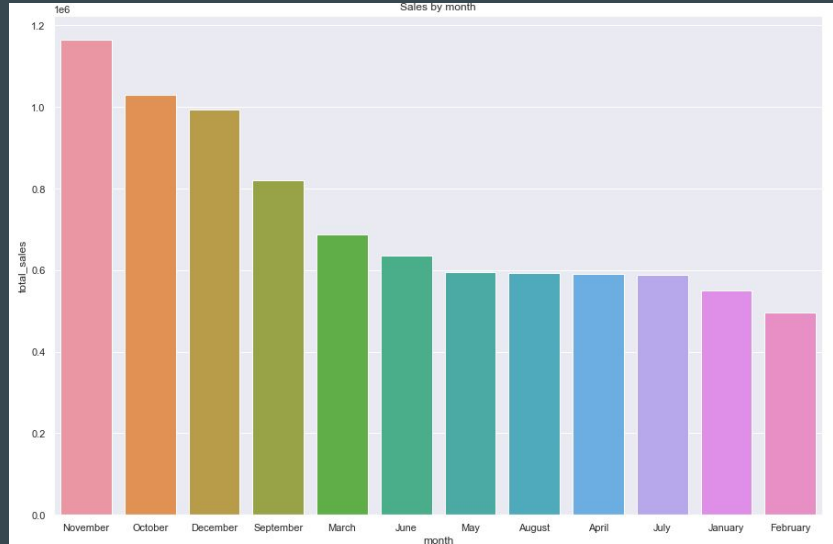
Portfolio Project

Exploratory Data Analysis - Analyzing Total Sales per Year



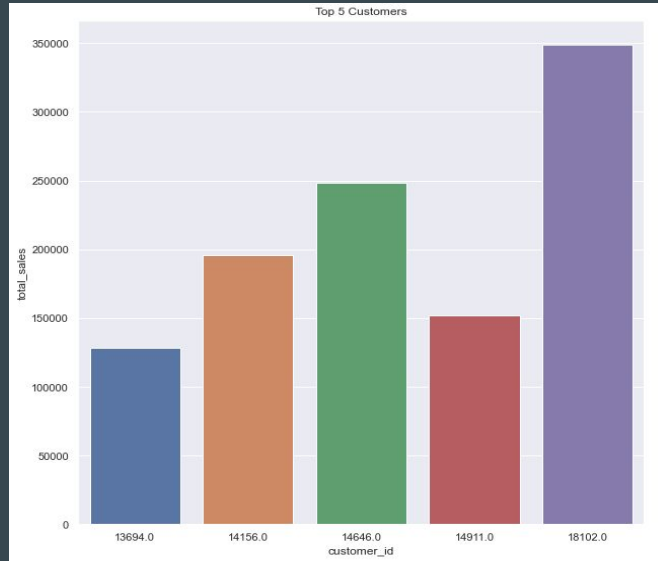
Portfolio Project

Exploratory Data Analysis - Analyzing Total Sales by Month



Portfolio Project

Exploratory Data Analysis - Top 5 Customers by Total Sales



Portfolio Project

Feature Engineering

- Total Sales (Quantity * Price)
- Extracted Year, Month, Day & Hour
- Created Bins by Total Sales

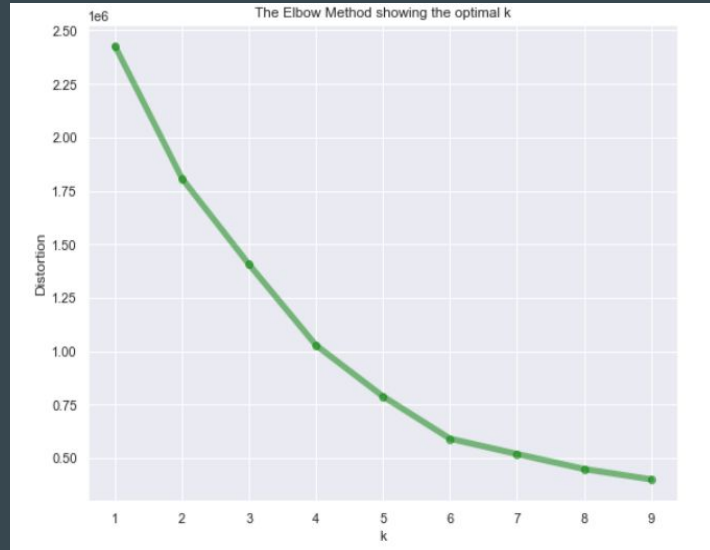
Portfolio Project

K-Means Clustering

- Capable of clustering quickly and efficiently
- Suitable when you have fewer features
- Simplicity of implementation
- Adapts to new examples
- Widely used in the Retail Industry

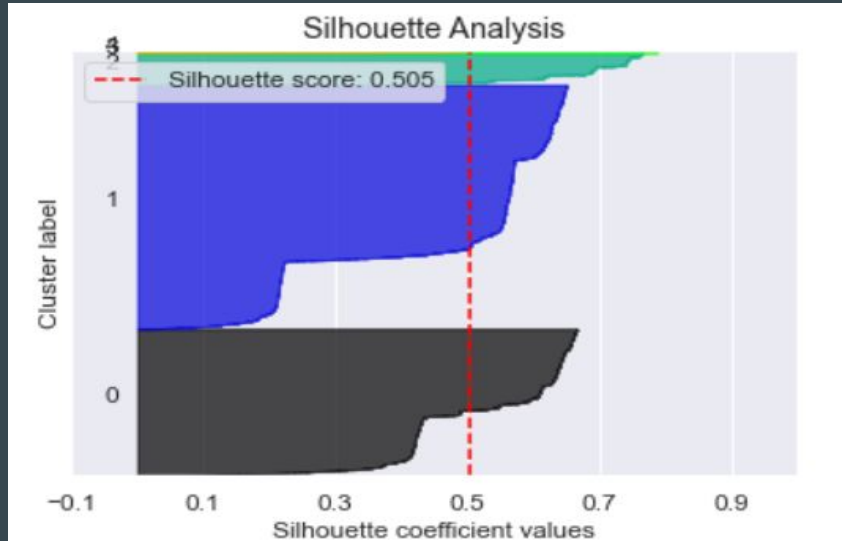
Portfolio Project

K-Means Clustering: Selecting clusters (Elbow Method)



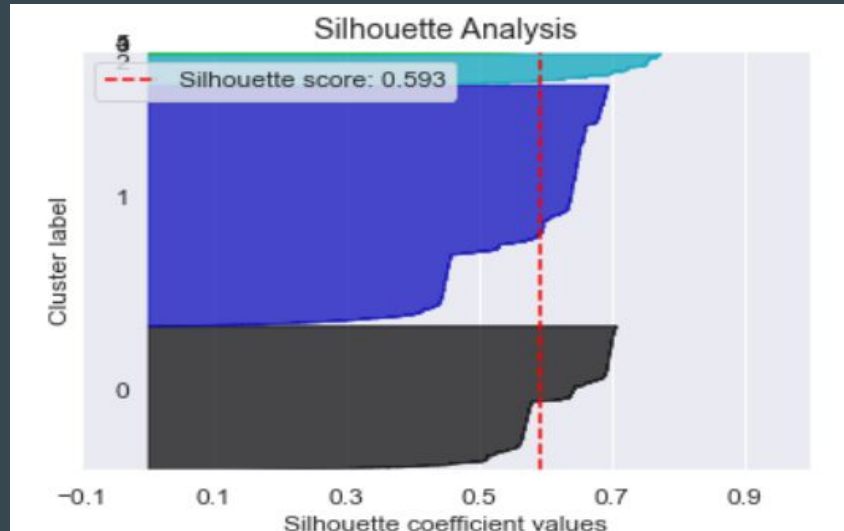
Portfolio Project

K-Means Clustering: Selecting clusters (Silhouette Analysis, k=5)



Portfolio Project

K-Means Clustering: Selecting clusters (Silhouette Analysis, k=6)



Portfolio Project

K-Means Clustering: Silhouette score observations

- Computationally expensive
- For most clusters, Silhouette scores were above 0.50

Portfolio Project

K-Means Clustering: Analyzing the clusters

- Although Elbow Method & Silhouette scores suggest choosing five or six clusters, it was more logical to have three clusters instead

Portfolio Project

K-Means Clustering: Analyzing the clusters

Cluster # 2

invoice	inf
quantity	1.033452e+01
price	3.158353e+00
customer_id	1.547881e+04
total_sales	1.870086e+01
year	2.009000e+03
hour	1.310505e+01
spend_category	0.000000e+00
cluster	2.000000e+00

Cluster # 3

invoice	9.922310e+28
quantity	1.000000e+00
price	8.864848e+03
customer_id	1.529000e+04
total_sales	8.864848e+03
year	2.010000e+03
hour	1.280000e+01
spend_category	1.000000e+00
cluster	3.000000e+00

Cluster # 4

invoice	inf
quantity	2.913678e+02
price	2.647586e+01
customer_id	1.547551e+04
total_sales	5.669323e+02
year	2.009926e+03
hour	1.215449e+01
spend_category	1.000000e+00
cluster	4.000000e+00

Portfolio Project

Deep Learning

Used Deep Learning for classifying the clusters that I generated

- Sequential model
- Two hidden layers (100 neurons each)
- Softmax (for multiclass classification)
- Sparse categorical cross entropy for loss
- ReLu for Activation
- Accuracy for evaluation

Portfolio Project

Next steps

Accuracy score of the Deep Learning Model: 0.5785

The current accuracy score is not at the desired level, but it can serve as a starting point for further improvements in accuracy

Portfolio Project

Conclusion

- ❑ Successfully implemented K-means Clustering & Deep Learning techniques for customer segmentation
- ❑ The project sharpened my skills in data exploration, feature engineering, model optimization & interpretation
- ❑ Achieved a baseline accuracy of 0.5785, providing a starting point for future improvement and exploration of advanced techniques
- ❑ Laid a solid foundation for further exploration of advanced topics in Data Science and data-driven decision-making

Portfolio Project

Thank you for your attention!