Berkeley Engineering | Berkeley Haas Professional Certificate in Machine Learning and Artificial Intelligence

HANDE GABRALI-KNOBLOCH

16.05.2023

- Executive Summary
- Introduction
- Methodology
- Results
- Discussion
- Conclusion
- Appendix

EXECUTIVE SUMMARY

Summary of Methodologies

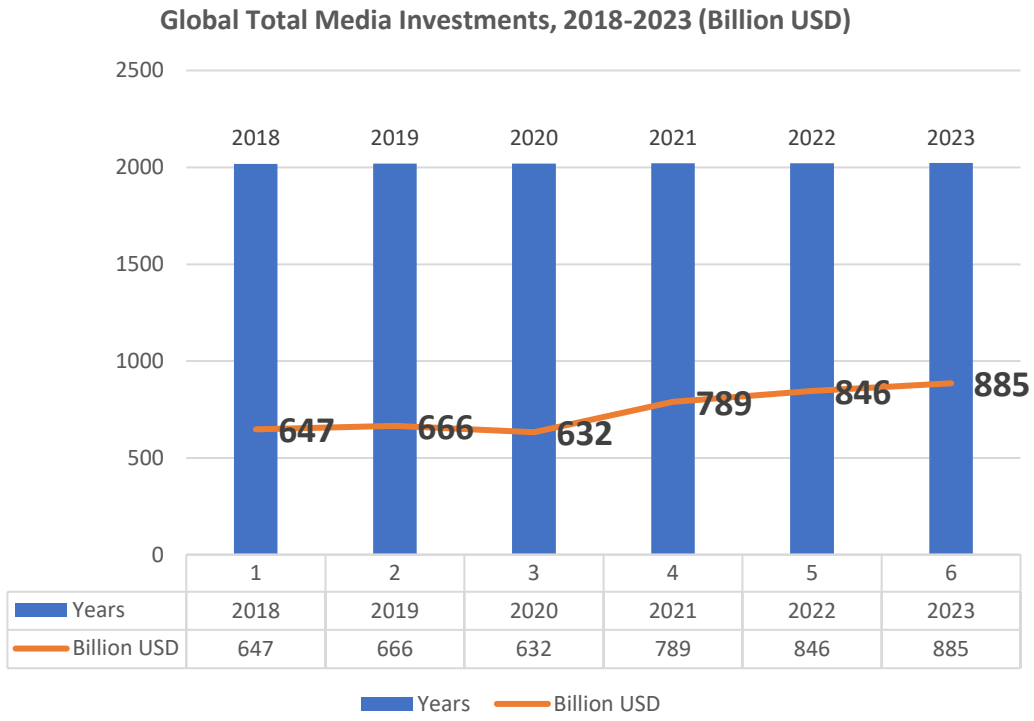
1. Data Collection Methodology:
2. Data Wrangling:
   - Data Cleaning
   - Data Standardization, Feature Scaling
3. Exploratory Data Analysis (EDA) using Visualization
4. Predictive Analysis using Supervised & Unsupervised Learnings
5. Modul Evaluation

INTRODUCTION

Project Background and Context

- With the effect of the Covid epidemic, consumers' staying at home for a long time has also changed their media usage habits.
- While the use of digital media in developed countries is progressing very rapidly, it is not possible to talk about the same speed in developing countries.
- The Coca-Cola Company is 137 years old in the world trade market with 200 brands worldwide. The innovations and changes it has brought to the market, it operates in approximately 200 countries globally.
- The Coca-Cola shapes its investments in the countries where it is active, as in every brand today, where the use of digital media surpasses traditional media.
- The social media and global video streaming services can potentially reshape consumers purchasing behaviors and media usage habits. Therefore, the media has an unavoidable function in conveying information and messages to the community.
- There are two leading media: traditional media, printed newspapers, magazines, billboards, books, brochures, television, radio, and others. Second is social or cyber media such as newspapers, blogs, Twitter, Instagram, and others. Many factors influence consumer intention toward brands and their purchase intentions (Noor, at., 2023, p.15).
- This capstone project also starts from the example of Turkey, which will be tried to reveal how the global brand Coca-Cola, particularly traditional media investments, are shaped in developing countries and their markets.

INTRODUCTION



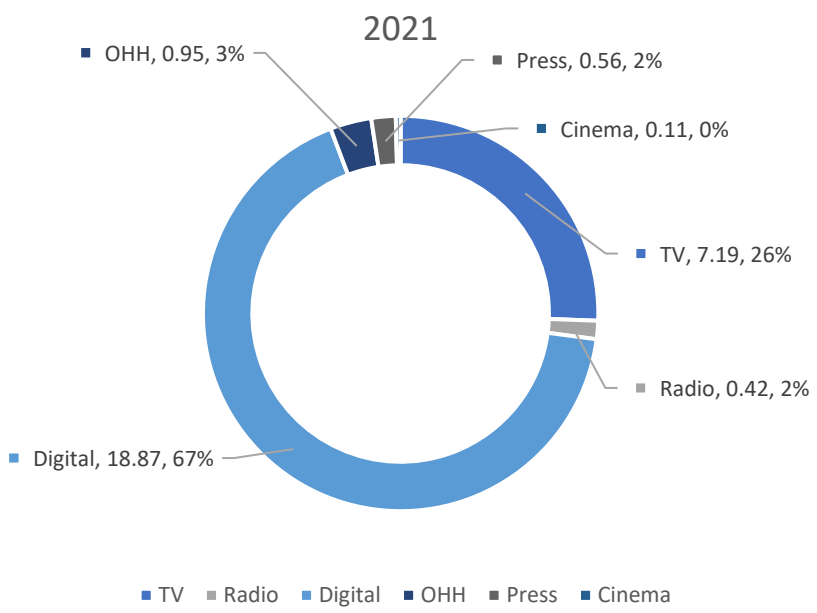
Media investments in the World:

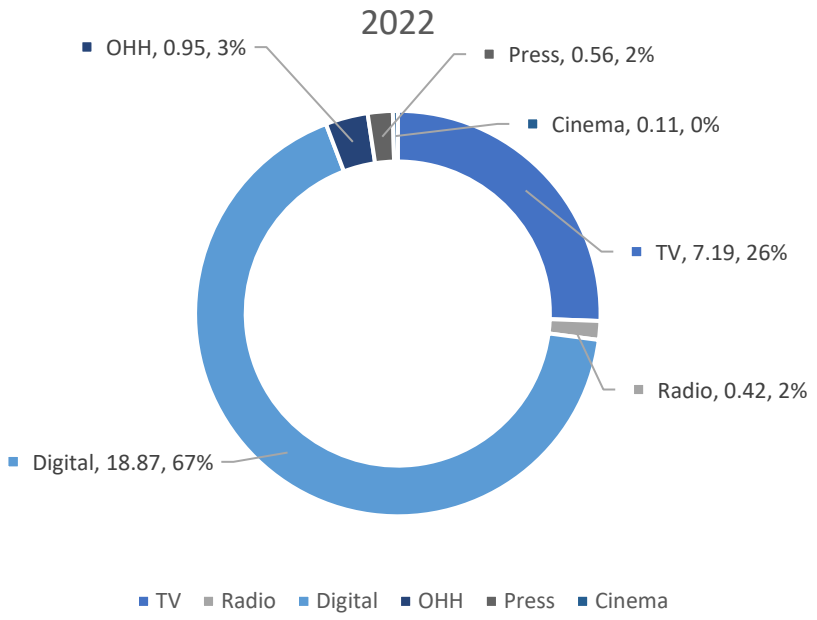
- Total media investments in the world are increasing
- continuously.
- 2020 as the effects of the pandemic diminish
- advertising with 24.7% growth after the year
- investments compared to previous years.
- Looking at the year 2022, the investment made
- continues to increase and worldwide
- Total media estimated to grow 7.31%
- Investments will grow by 4.55% next year (2023).

Resource: https://rd.org.tr/Assets/uploads/7587437b-563d-4917-b767-676021317bb1.pdf

INTRODUCTION

Media investments in Turkey 2021-2022
(Total organized media and direct purchasing channels, billion TL)





Resource: https://rd.org.tr/Assets/uploads/7587437b-563d-4917-b767-676021317bb1.pdf

INTRODUCTION

Problem Statement

In Turkey;

- TV advertising investments are with 78% increased in 2022 (fastest annual growth rate reached).
- Media investments in 2022, 24.03% in newspaper while the increase on the magazine side was 31.03%. took place.
- Outdoors closed 2022 with record growth: 124.63%.
- In 2022, radio investments exceeded 818 million TL. It has grown by nearly 100%.
- Cinema investments took place 110.7 Million TL in 2022.

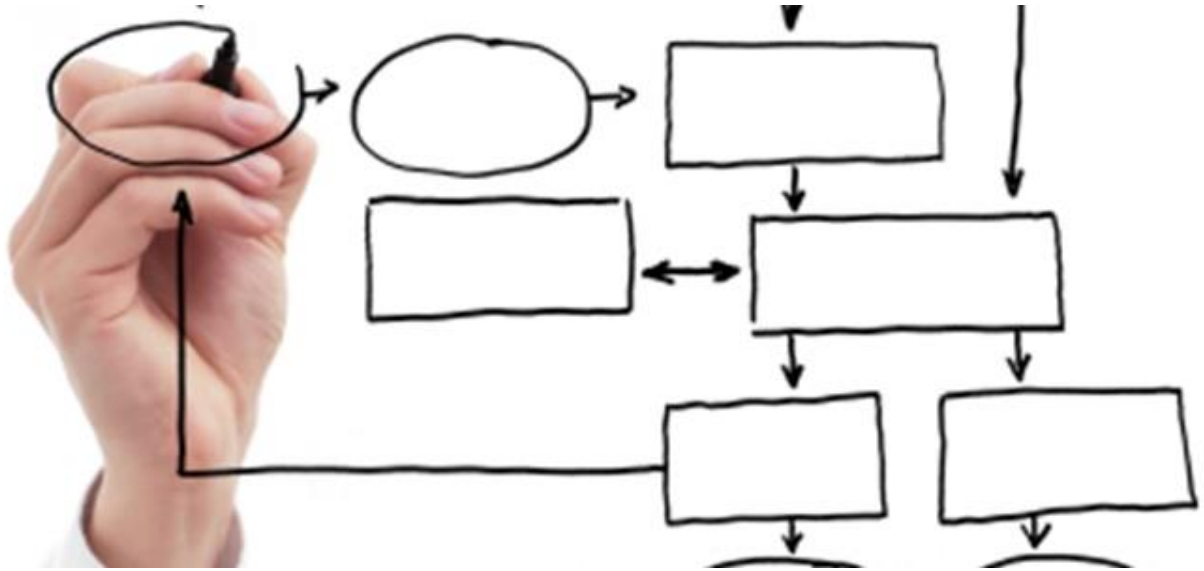
- Considering the growth in the traditional media industry in Turkey, The Coca-Cola company decided to conduct research on the setups of the channels to determine which variables in the campaign had an impact on traditional media and greater impact on the performance and outcome of the campaign.
- Thus, starting from the total media investments in Turkey in 2020, it is desired to evaluate the general investment tables of 2022 and get an idea about 2023 and beyond. As a result, the consistency of the analyzes made regularly by the brand will also be reviewed.

Common Problems that are needed to be solved

- What affects the traditional media investments and communications of brands?
- Why should the Coca-Cola company continue to invest in mainstream media?
- Does the usage of TV's GRP positively affects the TV investment of brands?
- What do the traditional media investments costs depend on?
- What and how is it determined (whether or not) the first stage will be used again?
- Under what conditions can the traditional media investments rates be determined?

Resource: https://rd.org.tr/Assets/uploads/7587437b-563d-4917-b767-676021317bb1.pdf

METHODOLOGY

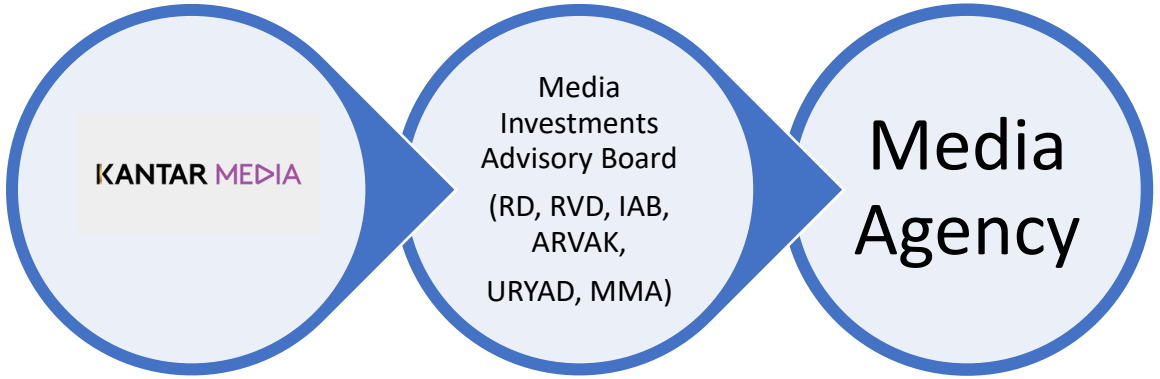


Summary of Methodologies

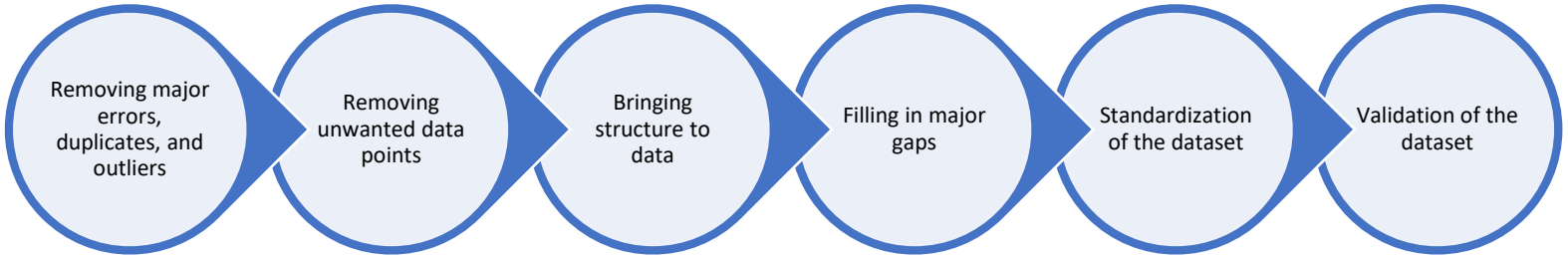
1. Data Collection Methodology:
2. Data Wrangling:
- Data Cleaning
- Data Standardization, Feature Scaling
- 3. Exploratory Data Analysis (EDA) using Visualization
- 4. Predictive Analysis using Supervised & Unsupervised Learnings
- 5. Modul Evaluation

Data Collection Methodology

- Dataset is a second-party data which is the first-party data of other organizations.
- **Dataset:**
- Based on the problem statement and questions, the following dataset may be used in this research:
- **Traditional/ Mainstream advertising data from 2020:** This dataset would provide information on the performance of mainstream campaigns during the Corona pandemic. It would include metrics such as GRP rates (Gross Rating Point) , day parts, frequencies, Column x cm (Press), item count, and costs.
- Besides, this dataset would provide information on the performance of NARDT market campaigns during different months throughout the year.
- **Demographic data:** The target audiences used by the Coca-Cola Company in its campaigns and purchases are not included in the dataset due to the confidentiality of purchasing.
- Before all applications, raw dataset has 28.869 rows x 30 columns.



Data Wrangling



1. Data Cleaning
2. Data Standardization, Feature Scaling
3. Data Reduction
4. Variable Transformation

Missing Values

```
[103]: df.isnull().values.any()  # Are there any missing observations (values)?
```

```
[103]: False
```

```
[104]: df.isnull().sum() # How many missing observations in which variable?
```

```
[104]: Medium Type           0
       Channel Type          0
       Source Company        0
       Source Subsector      0
       Report Product        0
       Report Sector         0
       Report Subsector      0
       Report Brand          0
       Month                 0
       Year                  0
       Spot Type             0
       Daypart               0
       Report Custom 3       0
       Item Count            0
       Seconds               0
       Stcm_Press            0
       Page_Press            0
       Frequency _Outdoor    0
       NEW Cost_TL           0
       Cost_TL               0
       GRP_TOTAL             0
       30"GRP_TOTAL          0
       dtype: int64
```

Changing some column types to categories: Objects to categorical

```
[8]: df[df.select_dtypes(['object']).columns] = df.select_dtypes(['object']).apply(lambda x: x.astype('category'))
```

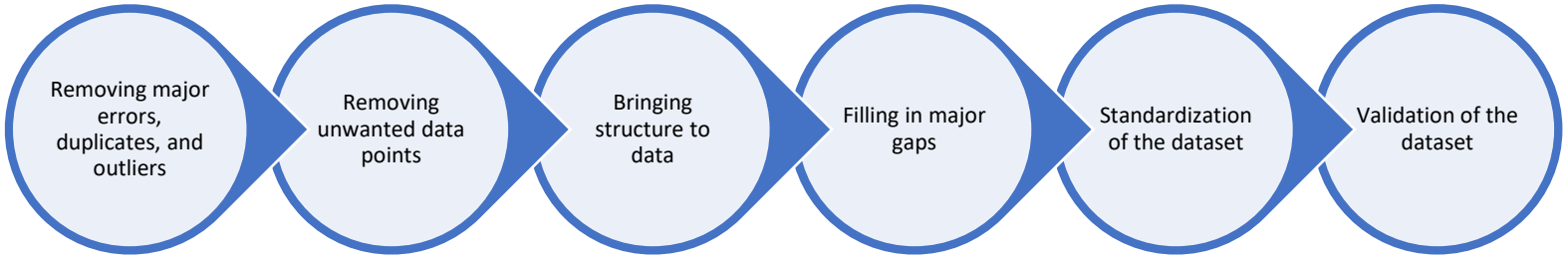
```
[9]: df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 28869 entries, 0 to 28868
Data columns (total 22 columns):
 #   Column              Non-Null Count  Dtype
---  ------              --------------  -----
 0   Medium Type         28869 non-null  category
 1   Channel Type        28869 non-null  category
 2   Source Company      28869 non-null  category
 3   Source Subsector    28869 non-null  category
 4   Report Product      28869 non-null  category
 5   Report Sector       28869 non-null  category
 6   Report Subsector    28869 non-null  category
 7   Report Brand        28869 non-null  category
 8   Month               28869 non-null  int64
 9   Year                28869 non-null  int64
 10  Spot Type           28869 non-null  category
 11  Daypart             28869 non-null  category
 12  Report Custom 3     28869 non-null  category
 13  Item Count          28869 non-null  int64
 14  Seconds             28869 non-null  int64
 15  Stcm_Press          28869 non-null  float64
 16  Page_Press          28869 non-null  int64
 17  Frequency _Outdoor  28869 non-null  int64
 18  NEW Cost_TL         28869 non-null  float64
 19  Cost_TL             28869 non-null  float64
 20  GRP_TOTAL           28869 non-null  float64
 21  30"GRP_TOTAL        28869 non-null  float64
dtypes: category(11), float64(5), int64(6)
memory usage: 2.8 MB
```

Coca-Cola

BerkeleyHaas

Data Wrangling



1. Data Cleaning
2. Data Standardization, Feature Scaling
3. Data Reduction
4. Variable Transformation

Outliers

```
[10]: df_outliers = df.select_dtypes(include = ['float64', 'int64'])
      df_outliers = df.dropna()
      df_outliers.head()
```

[10]:

	Medium Type	Channel Type	Source Company	Source Subsector	Report Product	Report Sector	Report Subsector	Report Brand	Month	Year	...	Report Custom 3	Item Count	Seconds	Stcm_Press	Page_Press	Frequency _Outdoor	NEW Cost_TL	Cost_TL	GRP_TOTAL	30"GRP_TOTAL
0	CINEMA	NON-TV	COCA COLA CORP.	SOFT DRINKS-FRUIT JOICE	CAPPY DESTEK MEYVE SUYU	NARTD	FRUIT JUICE	CAPPY	1	2020	...	MAINTHRUST	190	3610	0.0	0	0	0.0	16245.0	0.0	0.0
1	CINEMA	NON-TV	COCA COLA CORP.	SOFT DRINKS-FRUIT JOICE	COCA COLA LIME	NARTD	CSD	COCA-COLA	1	2020	...	MAINTHRUST	389	5835	0.0	0	0	0.0	26257.5	0.0	0.0
2	CINEMA	NON-TV	COCA COLA CORP.	SOFT DRINKS-FRUIT JOICE	COCA COLA LIME	NARTD	CSD	COCA-COLA	1	2020	...	MAINTHRUST	524	7860	0.0	0	0	0.0	35370.0	0.0	0.0
3	CINEMA	NON-TV	COCA COLA CORP.	SOFT DRINKS-FRUIT JOICE	COCA COLA LIME	NARTD	CSD	COCA-COLA	1	2020	...	MAINTHRUST	440	6600	0.0	0	0	0.0	29700.0	0.0	0.0
4	CINEMA	NON-TV	COCA COLA CORP.	SOFT DRINKS-FRUIT JOICE	COCA COLA	NARTD	CSD	COCA-COLA	1	2020	...	MAINTHRUST	2535	53240	0.0	0	0	0.0	239580.0	0.0	0.0

5 rows × 22 columns

Coca-Cola

BerkeleyHaas

Data Standardization, Feature Scaling

LabelEncoder

[112]:
```
from sklearn.preprocessing import LabelEncoder
lbe = LabelEncoder()
```

[113]:
```
lbe.fit_transform(df['Medium Type'])
```

[113]:
```
array([0, 0, 0, ..., 5, 5, 5])
```

[114]:
```
df['Medium Type']=lbe.fit_transform(df['Medium Type'])
```

[115]:
```
def label_encoder(dataframe, binary_col):
    labelencoder = LabelEncoder()
    dataframe[binary_col] = labelencoder.fit_transform(dataframe[binary_col])
    return dataframe
```

[116]:
```
binary_cols=['Medium Type','Channel Type','Report Product','Report Sector', 'Source Company ','Spot Type', 'Source Subsector ','Report Subsector','Report Brand','Report Custom 3','Daypart']

for col in binary_cols:
    label_encoder(df, col)
```

[117]:
```
df
```

[117]:

	Medium Type	Channel Type	Source Company	Source Subsector	Report Product	Report Sector	Report Subsector	Report Brand	Month	Year	...	Report Custom 3	Item Count	Seconds	Stcm_Press	Page_Press	Frequency _Outdoor	NEW Cost_TL	Cost_TL	GRP_TOTAL	30"GRP_TOTAL
0	0	1	15	4	19	0	4	14	1	2020	...	0	190	3610	0.0	0	0	0.000000	16245.000000	0.0	0.0
1	0	1	15	4	27	0	2	17	1	2020	...	0	389	5835	0.0	0	0	0.000000	26257.500000	0.0	0.0
2	0	1	15	4	27	0	2	17	1	2020	...	0	524	7860	0.0	0	0	0.000000	35370.000000	0.0	0.0

One-Hot Encoding

[118]:
```
df.head()
```

[118]:

	Medium Type	Channel Type	Source Company	Source Subsector	Report Product	Report Sector	Report Subsector	Report Brand	Month	Year	...	Report Custom 3	Item Count	Seconds	Stcm_Press	Page_Press	Frequency _Outdoor	NEW Cost_TL	Cost_TL	GRP_TOTAL	30"GRP_TOTAL
0	0	1	15	4	19	0	4	14	1	2020	...	0	190	3610	0.0	0	0	0.0	16245.0	0.0	0.0
1	0	1	15	4	27	0	2	17	1	2020	...	0	389	5835	0.0	0	0	0.0	26257.5	0.0	0.0
2	0	1	15	4	27	0	2	17	1	2020	...	0	524	7860	0.0	0	0	0.0	35370.0	0.0	0.0
3	0	1	15	4	27	0	2	17	1	2020	...	0	440	6600	0.0	0	0	0.0	29700.0	0.0	0.0
4	0	1	15	4	24	0	2	17	1	2020	...	0	2535	53240	0.0	0	0	0.0	239580.0	0.0	0.0

5 rows × 22 columns

Coca-Cola

BerkeleyHaas

Data Standardization, Feature Scaling

Standardization

```
[223]: # Standardization

       from sklearn import preprocessing

       df_std= preprocessing.scale(df)
```

```
[224]: df_std
```

```
[224]: array([[-9.10557483,  0.35464079, -1.17609792, ...,  0.74967938,
               -0.51829774, -0.45658304],
              [-9.10557483,  0.35464079, -1.17609792, ...,  1.45644041,
               -0.51829774, -0.45658304],
              [-9.10557483,  0.35464079, -1.17609792, ...,  2.09967237,
               -0.51829774, -0.45658304],
              ...,
              [ 0.26953982,  1.47405089, -0.03943546, ..., -0.37937358,
               -0.51829774, -0.45658304],
              [ 0.26953982,  1.47405089, -0.03943546, ..., -0.37683242,
               -0.51829774, -0.45658304],
              [ 0.26953982,  1.47405089, -0.03943546, ..., -0.38149122,
               -0.51829774, -0.45658304]])
```

Normalization:

```
[227]: # Normalization:
       df_nor= preprocessing.normalize(df_std)

       df_nor
```

```
[227]: array([[-0.78045066,  0.03039672, -0.10080488, ...,  0.06425599,
               -0.04442397, -0.03913433],
              [-0.57710931,  0.02247705, -0.07454083, ...,  0.09230887,
               -0.0328496 , -0.02893813],
              [-0.46620379,  0.01815754, -0.060216  , ...,  0.10750285,
               -0.02653675, -0.02337697],
              ...,
              [ 0.07463118,  0.40814066, -0.01091903, ..., -0.10504236,
               -0.14350819, -0.1264204 ],
              [ 0.0746927 ,  0.40847707, -0.01092803, ..., -0.10442476,
               -0.14362648, -0.1265246 ],
              [ 0.07462368,  0.40809964, -0.01091794, ..., -0.10561808,
               -0.14349377, -0.12640769]])
```

Min-Max Conversion

```
[121]: # Min-Max Conversion

       scaler = preprocessing.MinMaxScaler(feature_range = (100,200))
```

```
[228]: scaler.fit_transform(df_std)
```

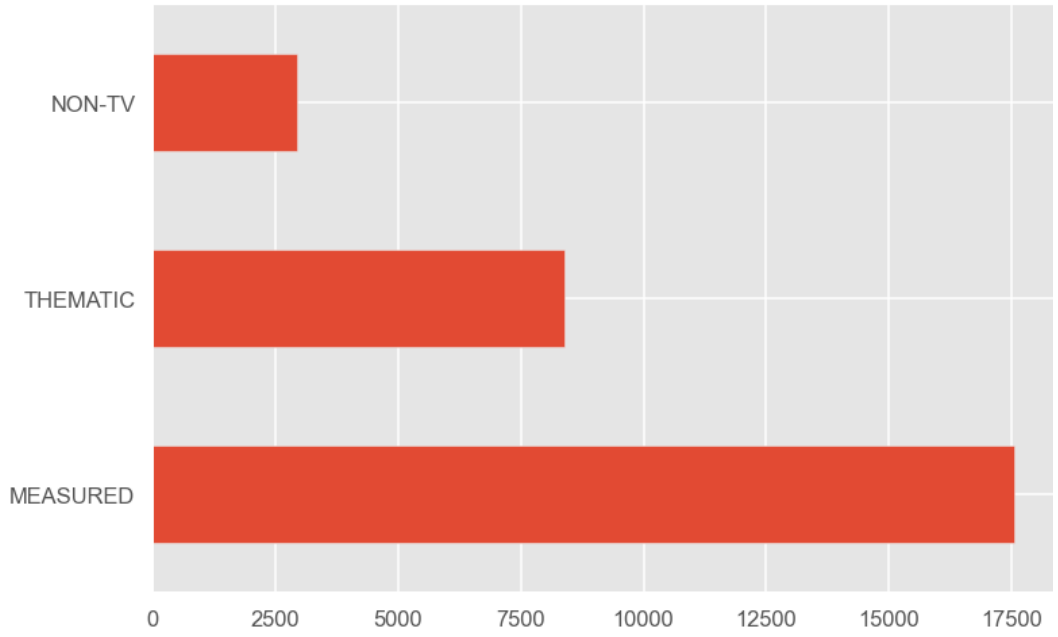
```
[228]: array([[100.        , 150.        , 120.27027027, ..., 104.275     ,
                100.        , 100.        ],
               [100.        , 150.        , 120.27027027, ..., 106.90986842,
                100.        , 100.        ],
               [100.        , 150.        , 120.27027027, ..., 109.30789474,
                100.        , 100.        ],
               ...,
               [200.        , 200.        , 152.7027027 , ..., 100.06578947,
                100.        , 100.        ],
               [200.        , 200.        , 152.7027027 , ..., 100.07526316,
                100.        , 100.        ],
               [200.        , 200.        , 152.7027027 , ..., 100.05789474,
                100.        , 100.        ]])
```

- After all processes, data has 28,869 entries and total 22 columns.

Exploratory Data Analysis (EDA) using Visualization

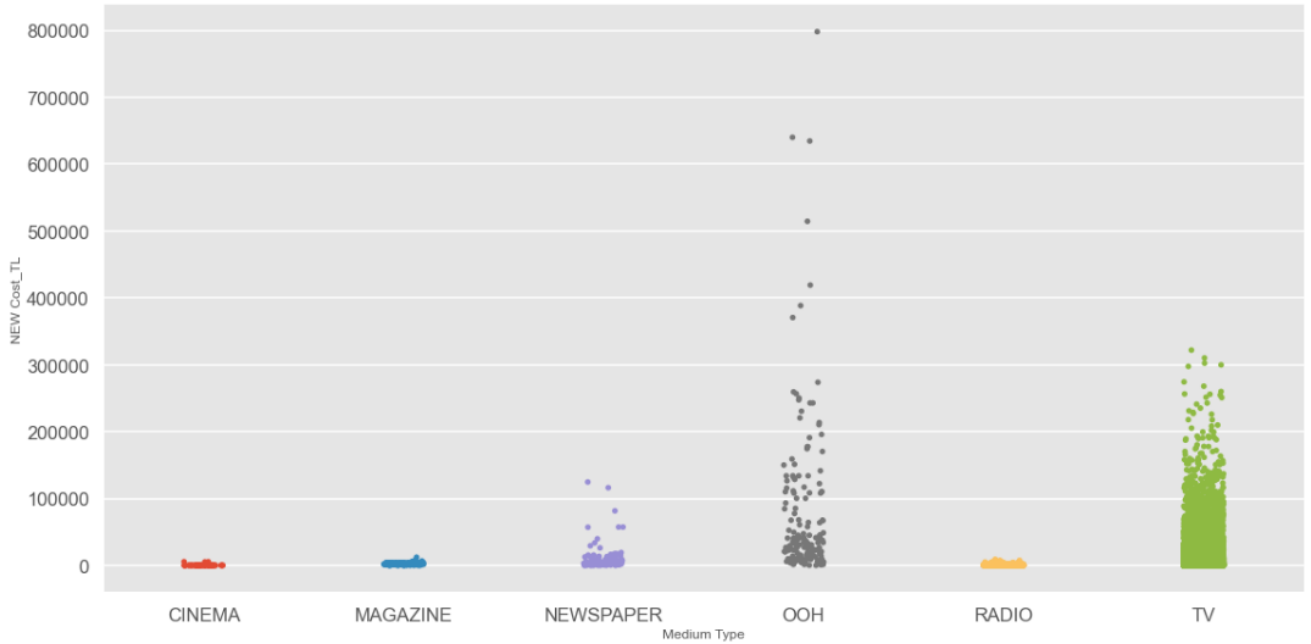
Bar Chart of Channel Type

```
[43]: df['Channel Type'].value_counts().plot.barh();
```



Categorical chart for Medium Type and New Cost

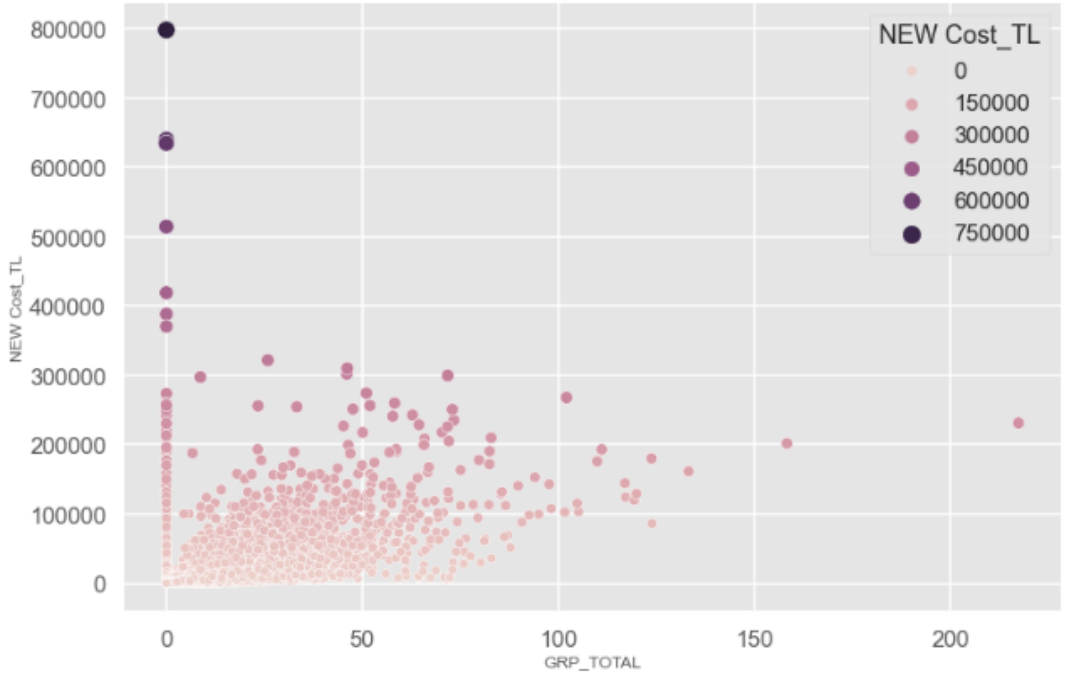
```
[104]: sns.catplot(x = 'Medium Type', y ='NEW Cost_TL', data = df, height=8, aspect=2,);
```



Exploratory Data Analysis (EDA) using Visualization

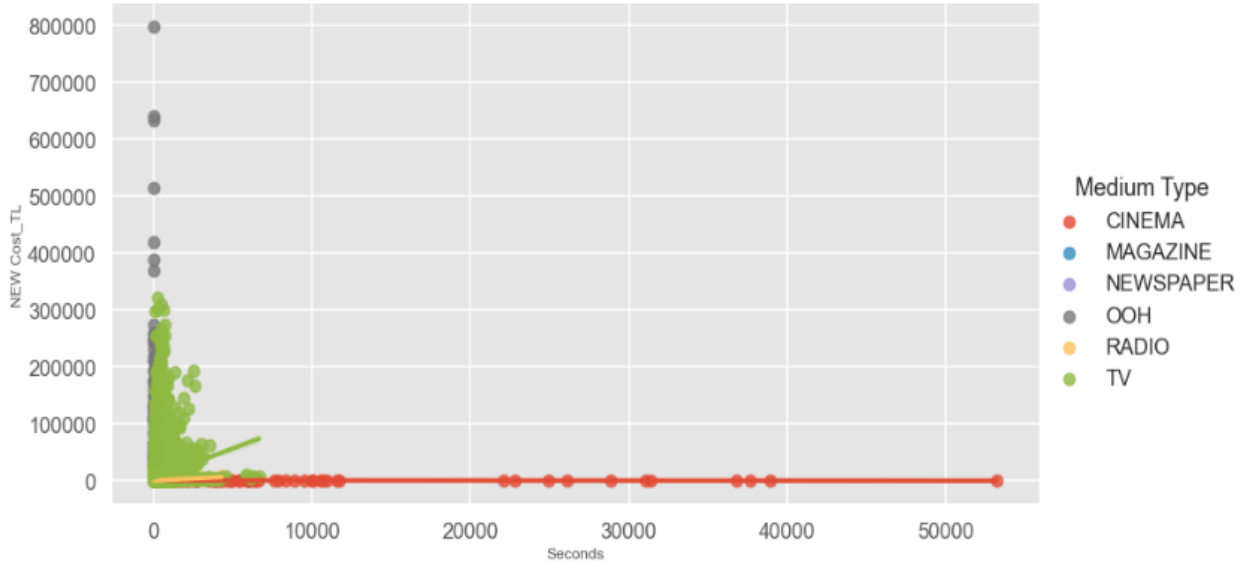
Scatterplot between GRP_TOTAL and New_Cost

```
[114]: sns.scatterplot(x = 'GRP_TOTAL', y ='NEW Cost_TL', hue='NEW Cost_TL', size = 'NEW Cost_TL', data = df);
```



LM Plot between Seconds used, NEw_Cost and Medium Type

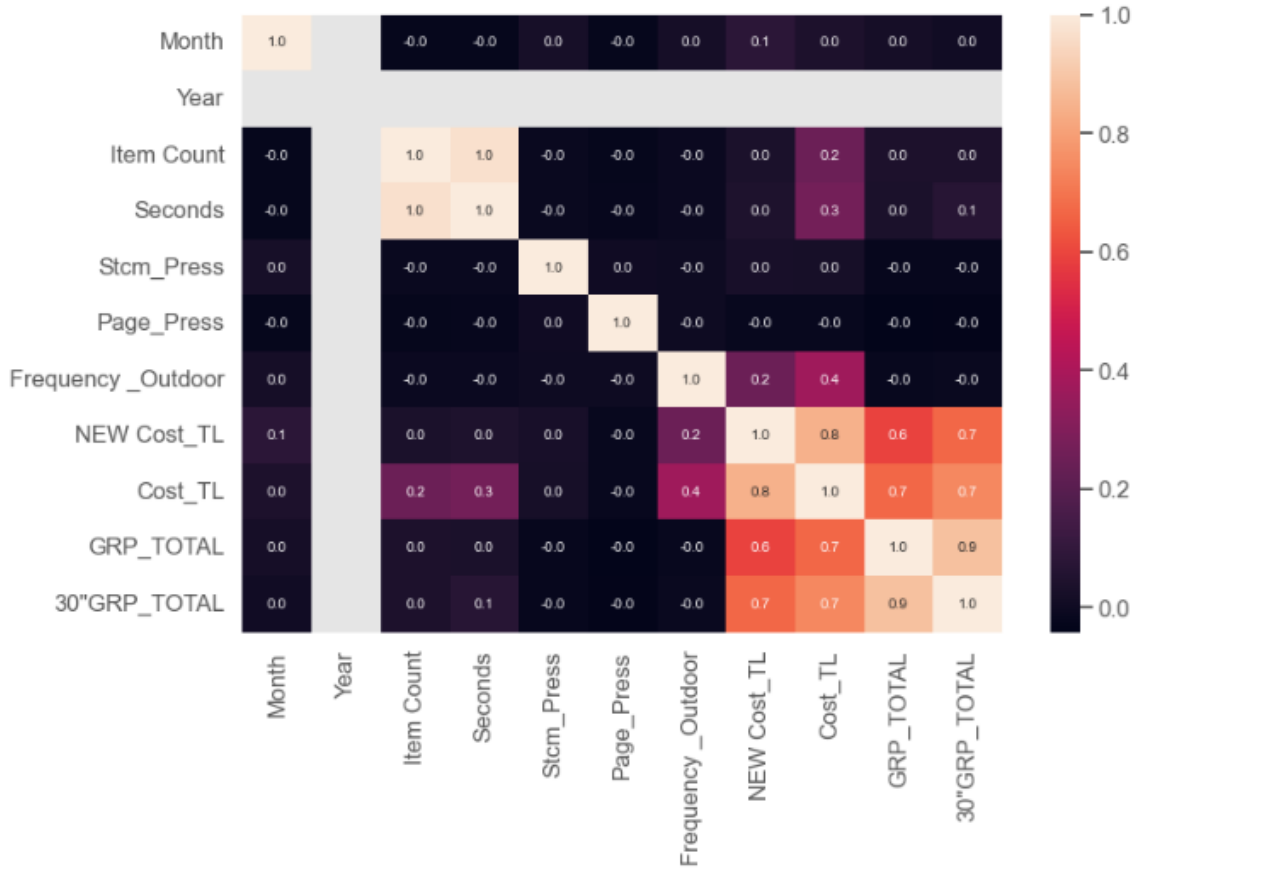
```
[122]: sns.lmplot(x = 'Seconds', y = 'NEW Cost_TL', hue = 'Medium Type', data = df,  height=6, aspect=2);
```



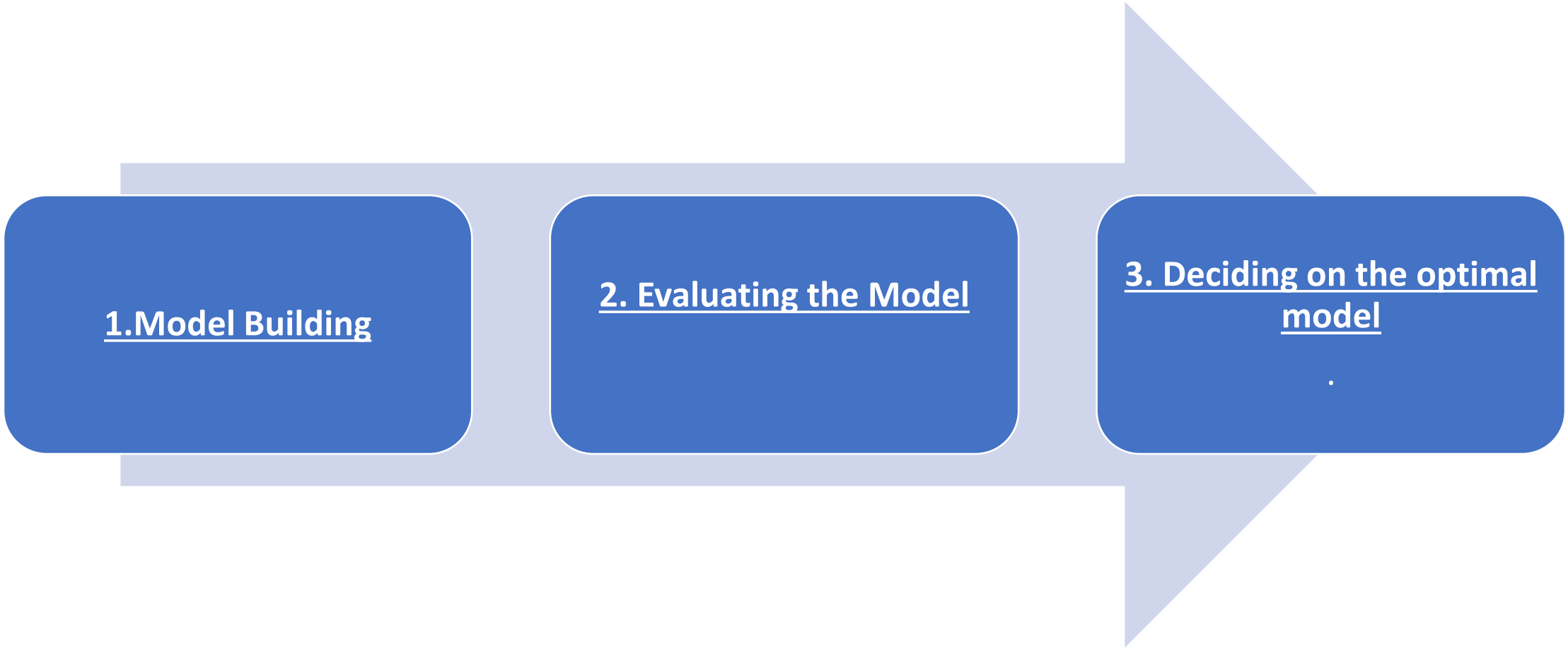
Exploratory Data Analysis (EDA) using Visualization

Heatmap

```
]: plt.subplots(figsize=(12,8))
   sns.heatmap(df.corr(), annot= True, fmt=".1f")
   plt.show()
```



Predictive Analysis using Supervised & Unsupervised Learnings



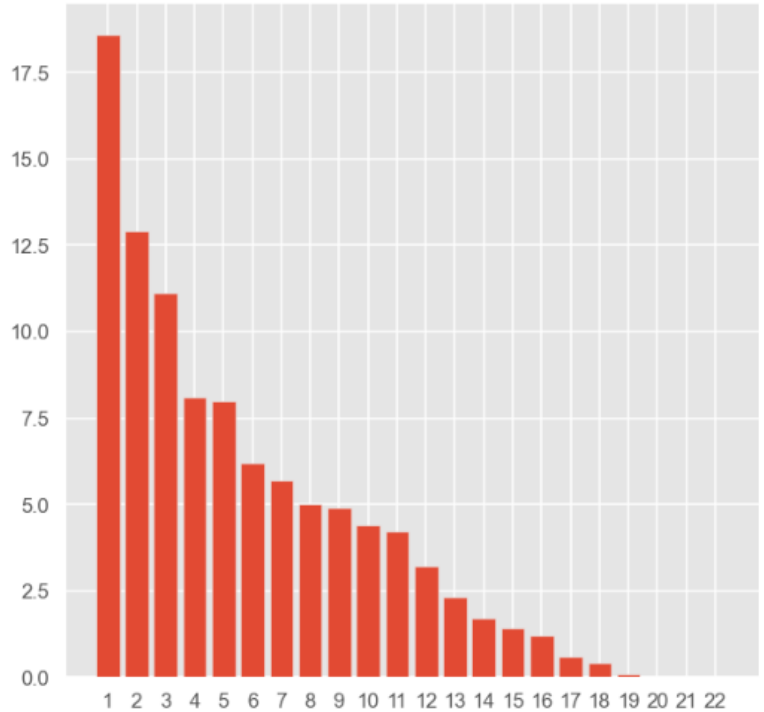
Predictive Analysis using Unsupervised Learning- Principal component analysis (PCA)

- These methods give us some methods of roughly how we can operate between independent variables when we don't have a variable at hand.

```
PCA Overview
=======================================
Total: 22 components
---------------------------------------
Mean explained variance: 0.045
---------------------------------------
    explained variance  cumulative
1             0.185959    0.185959
2             0.128770    0.314730
3             0.111189    0.425918
4             0.081061    0.506979
5             0.079533    0.586512
6             0.061828    0.648340
7             0.057299    0.705639
8             0.050164    0.755802
9             0.048902    0.804704
10            0.043844    0.848549
11            0.042258    0.890806
12            0.032055    0.922862
13            0.022760    0.945622
14            0.017135    0.962757
15            0.013780    0.976536
16            0.012418    0.988954
17            0.005625    0.994579
18            0.003932    0.998511
19            0.001360    0.999871
20            0.000129    1.000000
---------------------------------------
```

We plot and check the variance of the components:

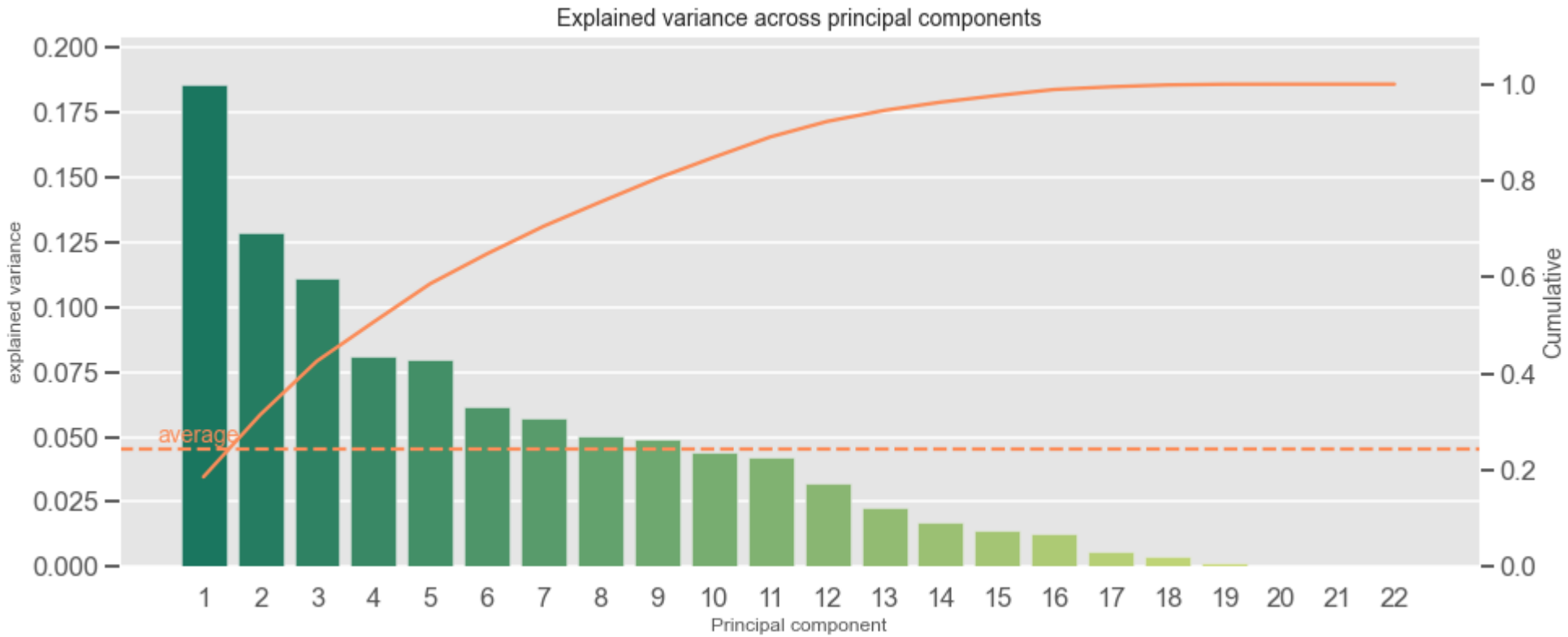
```
[96]:
plt.figure(figsize=(10,10))
var = np.round(pca.explained_variance_ratio_*100, decimals = 1)
lbls = [str(x) for x in range(1,len(var)+1)]
plt.bar(x=range(1,len(var)+1), height = var, tick_label = lbls)
plt.show()
```



- 4 is the variance drop-off point.
- The first three components explain the majority of the variance in our data.

Predictive Analysis using Unsupervised Learning- PCA

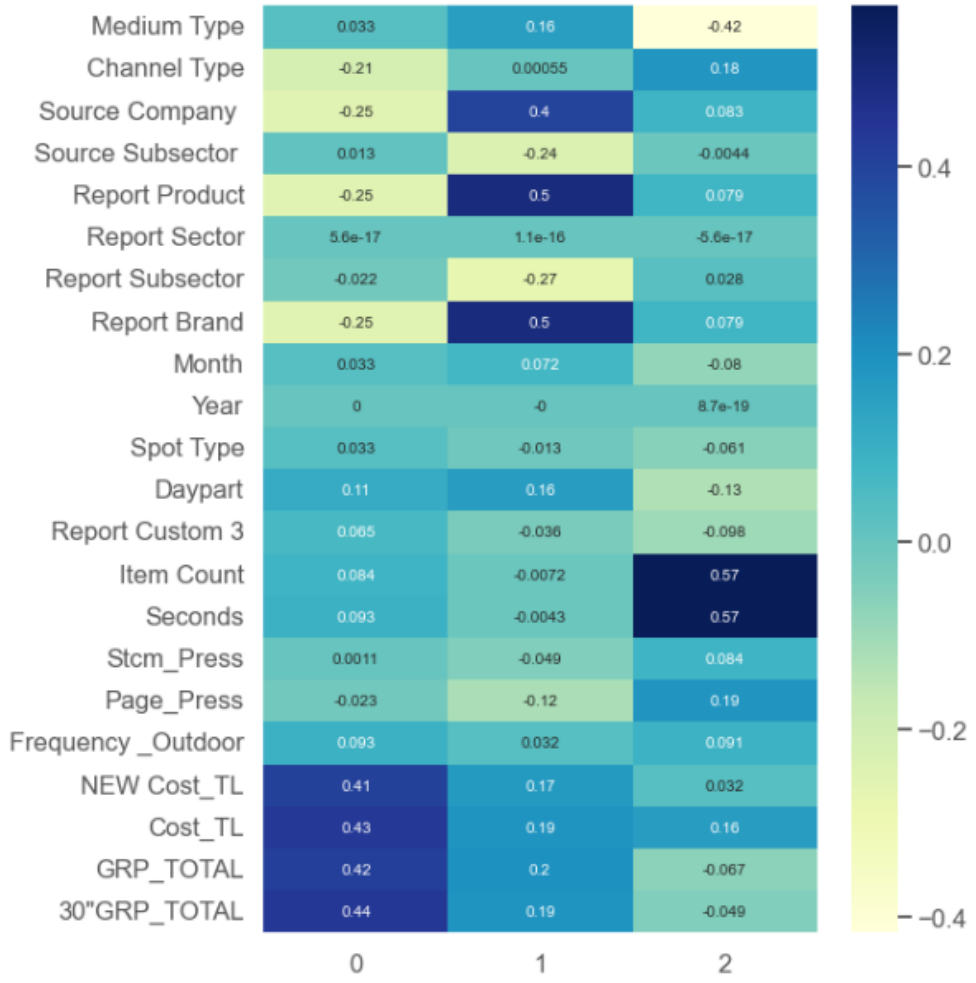
- Explained Variance across Principle Components:



Predictive Analysis using Unsupervised Learning- PCA

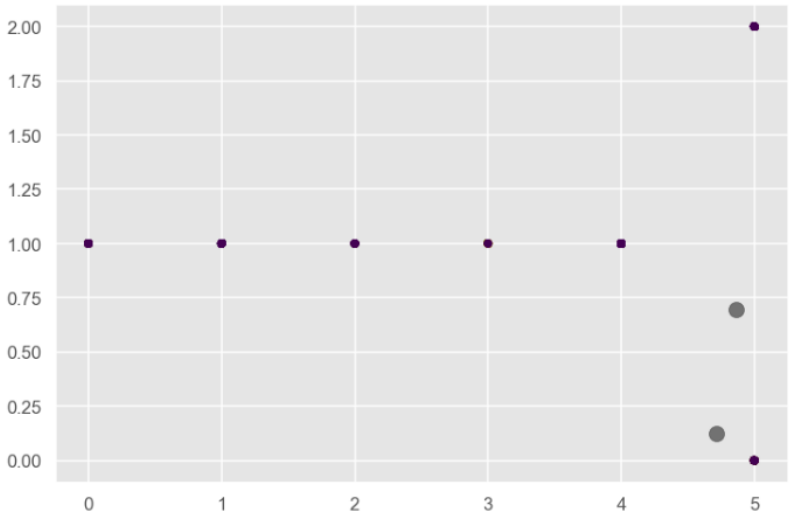
- Visualize factor loadings: Heatmap

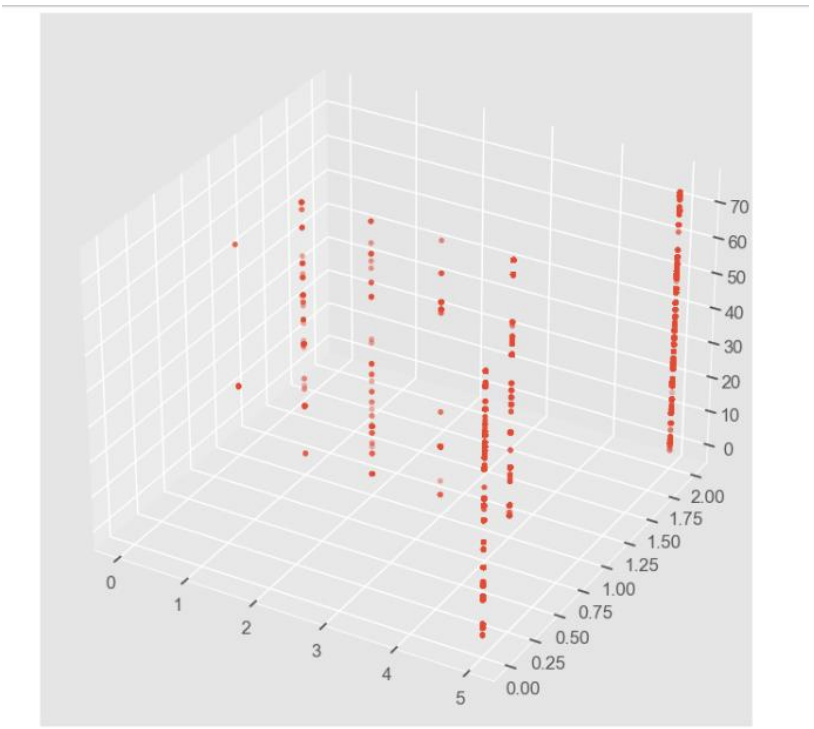
```
Factor Loadings for the 1. component
(explains 0.19 of the variance)
========================================

Top 3 highest
----------------------------------------
30"GRP_TOTAL    0.439178
Cost_TL         0.433636
GRP_TOTAL       0.416449
Name: 0, dtype: float64

Top 3 lowest
----------------------------------------
Source Company    -0.254392
Report Brand      -0.252849
Report Product    -0.248523
Name: 0, dtype: float64
```

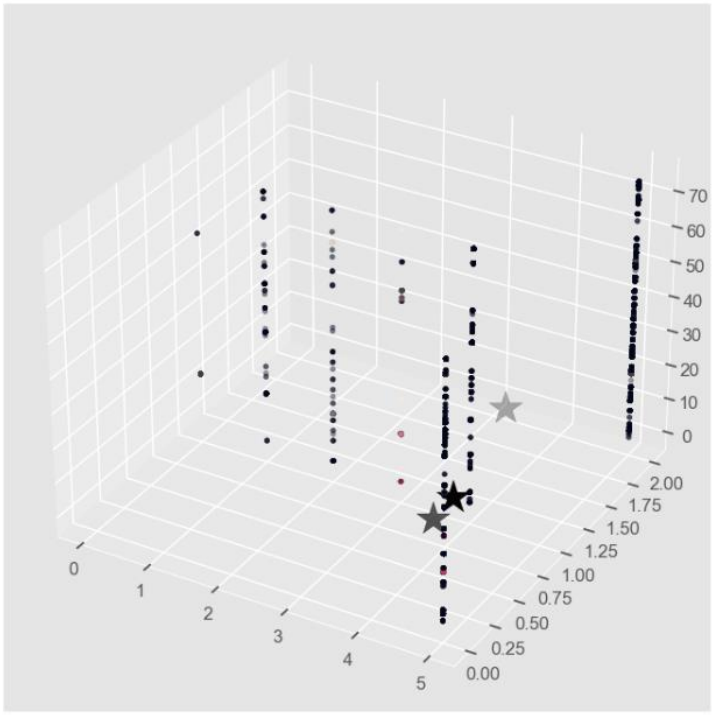


Predictive Analysis using Unsupervised Learning-K-Means

- This is unlabeled data and our objective is to find K number of groups or “clusters” which are similar to each other. Suppose our training set looks like this:







Coca-Cola

BerkeleyHaas

Predictive Analysis using Unsupervised Learning-K-Means

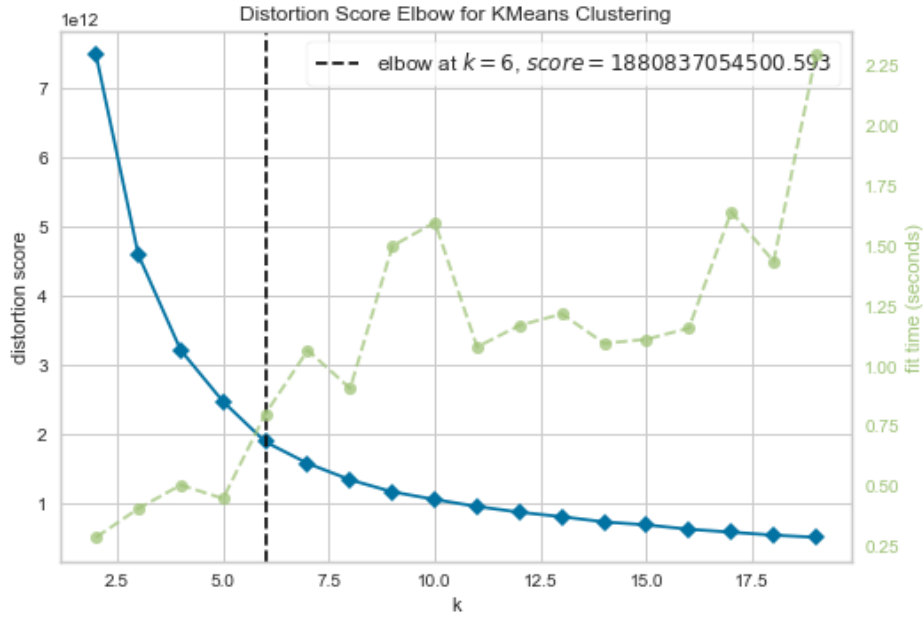
Clusters and Observation units

```
[140]: kmeans = KMeans(n_clusters = 3)
       k_fit = kmeans.fit(df)
       clusters = k_fit.labels_
```

```
[141]: pd.DataFrame({'Medium Type' : df.index, "Clusters": clusters})[0:10]
```

[141]:

	Medium Type	Clusters
0	0	0
1	1	0
2	2	0
3	3	0
4	4	1
5	5	2
6	6	2
7	7	2
8	8	0
9	9	0

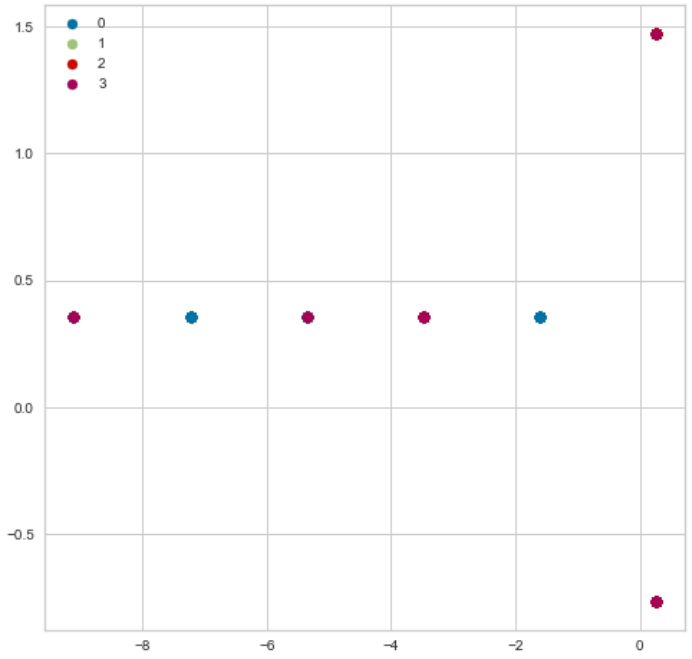



```
[144]: <AxesSubplot:title={'center':'Distortion Score Elbow for KMeans Clustering'}, xlabel='k', ylabel='distortion score'>
```

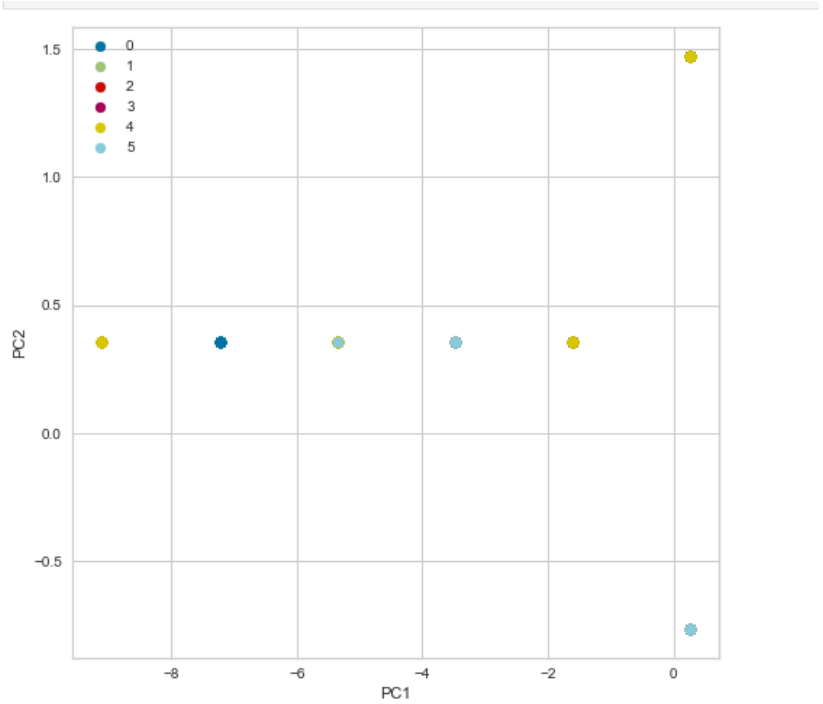
- **Determining the Optimum Number of Clusters:**
- In Model tuning part actually belongs to K-Means.
- It is desired that the similarity within clusters should be maximum and the similarity between clusters should be minimum.
- Mathematically, it is obtained by making and summing the distances of each observation unit from the center of the cluster it is in, and as a result, these calculations about the centers of all clusters in general.
- SSD: Here error refers to the distance.

Predictive Analysis using Unsupervised Learning-K-Means

- **Creating The Best Scoring Model:**
- "PCA helps to reduce the number of "features" while preserving the variance, whereas clustering reduces the number of "data-points" by summarizing several points by their expectations/means (in the case of k-means)“.
- **Training the models:**



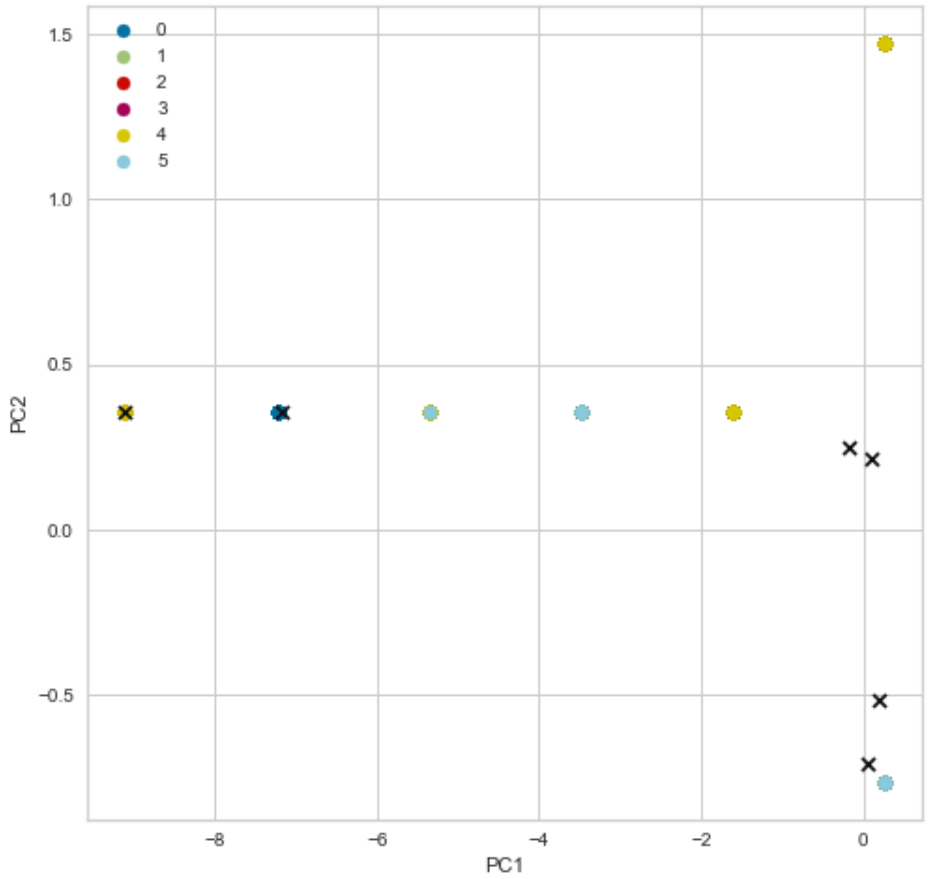
- With 4 clusters



- With 6 clusters

Predictive Analysis using Unsupervised Learning-K-Means

- Plotting centroids in 6 cluster model



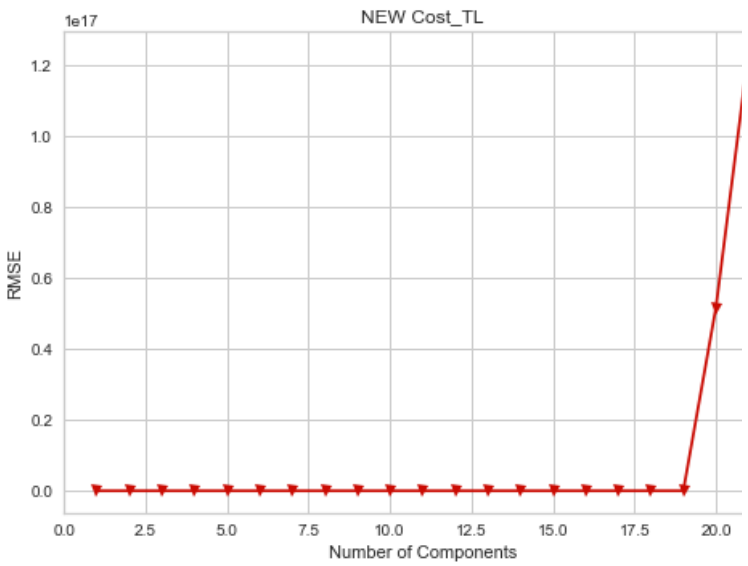
- Our goal here is to achieve high similarity within clusters and low similarity outside clusters.
- When we want to compartmentalize, K-means the most units of observation we have.
- The catch that will be encountered in K-Means is; It is a good understanding of its iterative nature. After calculating the distances of the initially created centers, the first primitive clustering structure occurs. We then recalculate the distances to a center.

Coca-Cola

BerkeleyHaas

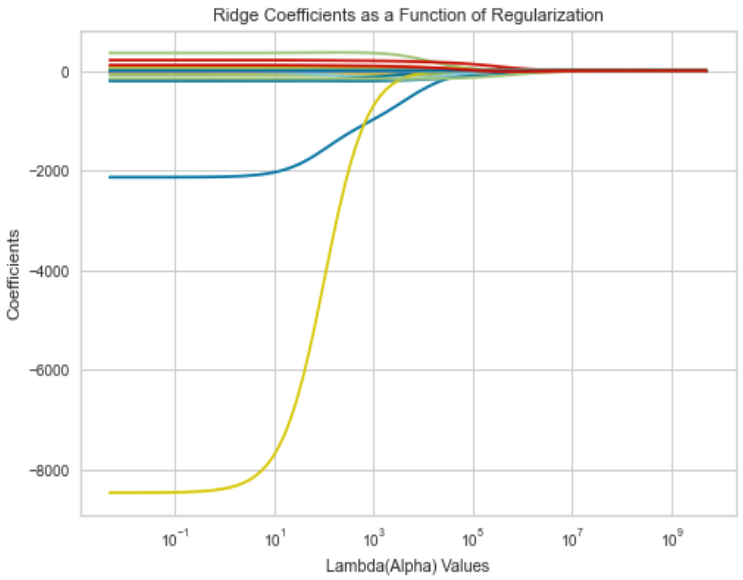
Predictive Analysis using Supervised Learning

PLS



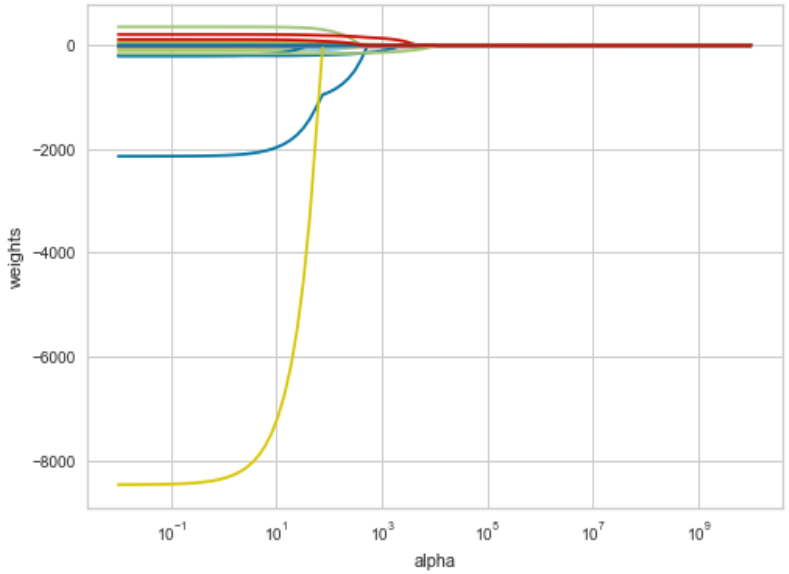
- 2 and 6 components and we have the information contained in these 22 variables.

RIDGE



- Purpose: To find the coefficients that minimize the sum of squares error by applying a penalty to these coefficients (Hoerl & Kennard, 1970).
- "Shows the effect of linearity in coefficients or Ridge. At the end of the path, the coefficients exhibit large oscillations as the alpha tends towards zero and the solution toward ordinary least squares".

LASSO



- Aim: To find the coefficients that minimize the sum of squares of error by applying a penalty to these coefficients (Tibshirani, 1996).
- Difference from Ridge: It makes variable selection by applying the penalties of the coefficients to make them zero.

Predictive Analysis using Supervised Learning- Modul Evaluation

Which model gives the best result for our dataset in general?			
Model Names	**Results**	**Cross_Val_Score**	
		Test	**Train**
Multiple Linear Regression		7549.517791	9288.241897
PLS	10974.56		
Ridge	9705.984		
Lasso	9709.243		

- Find the relevant and best hyperparameters for the model.
- Determine the best model that has the highest accuracy.
- Confirm the optimal model.

- To determine how good is a model, let us understand the impact of wrong predictions that are “mean_squared_error(y_test, y_pred)”. MLR, Ridge and Lasso are the best predicted models and the wors one is the PLS.

Coca-Cola

BerkeleyHaas

Summary of All Results

- In conclusion, data on TV, radio, press, cinema and OHH for 2020 were analyzed. The data was presented through various plots and charts to help understand the trends in frequencies, period of use, number of pages used for each medium types.
- While New_Cost TL is expressing the mainstream media investments made by the Coca-cola company in 2020, Multi linear regression analysis was used to determine the relationship between continuous variables in the dataset.
- To identify the strength of the effect that the independent variables have on a dependent variable. So, determining the strength of relationship between dose and effects of New_Cost TL and other variables. Also, we tried to forecast effects or impacts of changes.
- So cross_val_score.mean estimates the expected 76% accuracy of our model on out-of-training data. Based on the results, New_Cost TL in 2020 has statistically significant coefficients, indicating that ad cost has a positive effect on performance impressions for each medium types.
- Alpha (α) is the penalty term that denotes the amount of shrinkage (or constraint) that will be implemented in the equation So, Ridge Regression alpha= 0.005, Lasso Regression alpha equals to 0.11079288755091725.
- K-means tries to discover the least-squares partition of the data. PCA determines the least-squares cluster membership vector.
- Elbow curve at k=6 and score= 1880837054500.593 so that, our optimal number of clusters for K-means clustering is 6.

DISCUSSION

- Brand awareness, brand image, brand attitude, prosumers, lead users and purchase intention are the key role in the brand communication.
- Also, customers need information about the brand to develop their awareness. The buying decisions of consumers always influenced by advertising. Regardless, traditional media applications are seen as more reliable than social media for consumers today. Especially considering the information clustering in the digital environment.
- In order to increase brand awareness and awareness among consumers, brands should increase their frequency and increase their visibility with the periodic investments and campaigns they have made on the basis of media, as seen in their traditional media investments and communications.
- For these reasons, especially in developing economies and their markets, by continuing to invest in mainstream media, the Coca-Cola company can build brand awareness and consumer information in a more reliable legacy media landscape.
- It is seen that the use of TV's GRP has positive effects on TV investments of brands. It is noteworthy that The Coca-Cola company invests heavily in the channels measured in Turkey.
- The cost of traditional media investments may depend on the time and unit price of the sport used in TV, at the beginning the cost of the line x column, and on the purchase and rental of boards in OHH through agreements. Here again, the importance of media buying and strategies comes to the fore.
- In this study, it is not known to what extent it contributes to traditional media investments, since gender cannot be examined in the context of the target audience.


CONCLUSION

APPENDIX

Articles:

Noor., N, Puteh.,K, Nordin.,N, Amir., M, Amir., H, Sazali., F, Kamaluddin., M. " Effectiveness of Traditional Media and Social Media Sustainability Communication in Influencing Green Consumption Intention", Journal of Media and Information Warfare, Vol. 16 (1), 14-27, April 2023. Pp. 15-18.

Shao., C. (2023), "Changing Mass Media Consumption Patterns Before/After Relocation: East

Asian International Students' Mass Media Use and Acculturation Strategies", International Journal of Communication 17(2023), 1592–1612.

Websites:

The Coca-Cola Company: Access Date: 07.05.23, https://www.coca-colacompany.com/

Turkish Government: Access Date: 07.05.23, https://www5.tbmm.gov.tr/kanunlar/k6487.html#:~:text=MADDE%202%2D%208%2F6%2F,ve%20t%C3%BCketicilere%20y%C3%B6nelik%20tan%C4%B1t%C4%B1m%C4%B1%20yap%C4%B1lamaz.

Word Investment Report 2020: Access Date: 07.05.23, https://unctad.org/system/files/official-document/wir2020_overview_en.pdf

Turkish Advertising Association Reports: Access Date: 07.05.23,

https://rd.org.tr/assets/uploads/medya_yatirimlari_2019_.pdf

https://rd.org.tr/Assets/uploads/7587437b-563d-4917-b767-676021317bb1.pdf

https://rd.org.tr/Assets/uploads/6ac5fc0a-f3c9-4564-bb1f-102ddc579341.pdf

Dentsu ad spend report: Access Date: 07.05.23, https://www.dentsu.com/news-releases/dentsu-ad-spend-forecast-july-2022-release

Word Economic Forum: 07.05.23, https://www.weforum.org/agenda/2022/05/a-digital-silver-bullet-for-the-world/

https://www.insiderintelligence.com/forecasts/5d13a07a64fe7d034c2cc15a/5d139fb0b88aeb0b7c481d6c/

Statista: Digital Newspapers & Magazines – Worldwide: 07.05.23, https://www.statista.com/outlook/amo/media/newspapers-magazines/digital-newspapers-magazines/worldwide

APPENDIX

Websites:

Kantar: Access Date: 08.05.23, https://www.kantar.com/locations/turkey#_=

Traditional Media vs. New Media: Which is Beneficial: Access Date: 08.05.23

https://www.techfunnel.com/martech/traditional-media-vs-new-media-beneficial/

https://stats.stackexchange.com/questions/183236/what-is-the-relation-between-k-means-clustering-and-pca#:~:text=k%2Dmeans%20tries%20to%20find,means%20maximizing%20between%20cluster%20variance.

K-means Clustering via Principal Component Analysis: https://ranger.uta.edu/~chqding/papers/KmeansPCA1.pdf

https://github.com/barbie-in-tech/K-Means-Cluster/blob/main/K_Means_Cluster.ipynb

https://uc-r.github.io/kmeans_clustering

https://www.analyticsvidhya.com/blog/2020/07/types-of-feature-transformation-and-scaling/

APPENDIX

Television

Reported TV – RD member (Association of Advertisers) media planning and purchasing agencies in line with the Grpxsn times realized in the TV channels whose ratings are reported full-time based on estimates made by Estimates have been finalized by comparing with the commercial revenues declared to RTÜK (Türkiye Radio and Television Supreme Board).

Unreported TV – Realized ad seconds on other TV channels that do not measure full-time ratings but are reported within the scope of Kantar Media AdEx

Includes media investment estimates (excluding European channels). All spot and banner advertisements, advertorial, product placement, teleshopping and program support revenues, whose total advertisement time is measured, are included. Channels not covered by Kantar Media reporting are out of scope.

Press

All national and local newspapers, newspaper supplements, magazines and magazines that Kantar Media measures advertising areas are included.

Advertising investments that are not reported within the scope of Kantar AdEx, such as word ads, inserts, editorials, advertorials and events in predominantly local and sectoral press and classified ads.

Outdoor

It is based on the turnover reported by companies operating in the outdoor sector and the detection of possible anomalies as a result of comparison with previous years.

Digital outdoor figures include areas such as Led screen, shopping mall/airport/point of sale screens.

Large areas include areas such as walls, giantboards, parapets, glass surfaces.

Outdoor advertising units include areas such as CLP, billboard, megalight, megaboard, banner.

Other applications such as guerrilla applications, events, signage and promotional items are not covered.

Radio

It includes media investment estimates made in all national and local radio channels where Kantar Media measures the actual advertising time.

As of 2017, all spot and generation advertisements, product placement, tele-shopping and program support revenues, where the total advertising time is measured in media investment estimates

(sponsorship) and commercial communication revenues declared to the Radio and Television Supreme Council were also started to be used as a source.

Channels not covered by Kantar Media reporting are out of scope.

Local investments are included in the calculation by estimating in line with the opinions of local media agencies.

Cinema

All movie theaters where Kantar Media measures the advertising duration are included in the scope. Investment figures only include screen advertisements whose duration is reported,

It does not cover foyer areas in the cinema, racket use in the hall and sponsorships.

Coca-Cola

BerkeleyHaas