Data Analysis(Part-III)

🎯 Learning Goals:

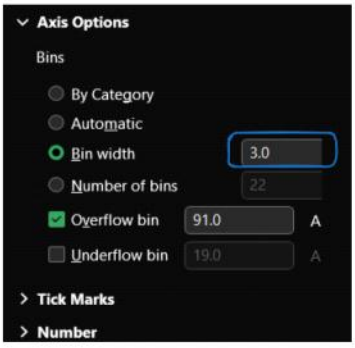
- Understand what bivariate analysis is
- Identify different types of bivariate analysis
- Perform bivariate analysis using PivotTables, correlation, scatter plots, and box plots
- Understand the importance of effective communication in analytics.
- Learn techniques to communicate insights and findings clearly.
- Be able to interpret and explain data visuals to stakeholders.

Correlation	Strength
0.00–0.19	Very Weak
0.20–0.39	Weak
0.40–0.59	Moderate
0.60–0.79	Strong
0.80–1.00	Very Strong

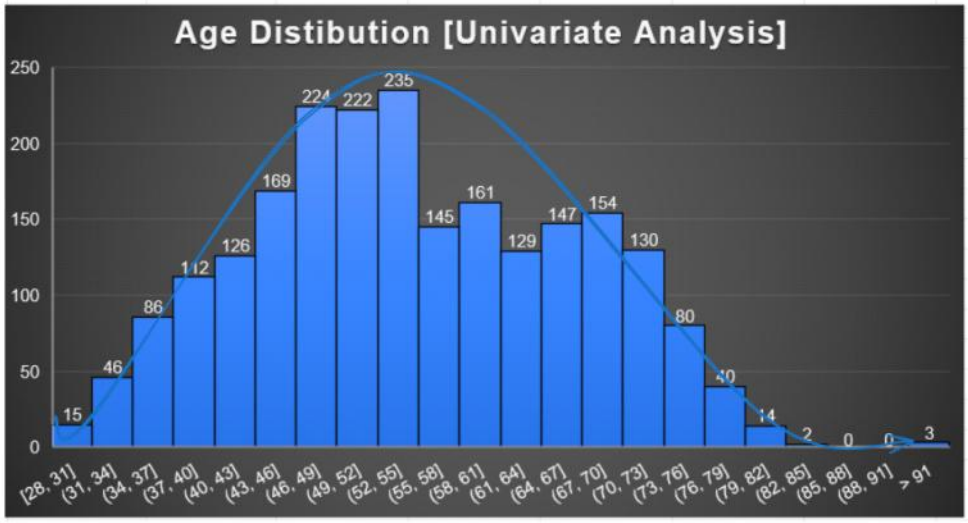
Histogram

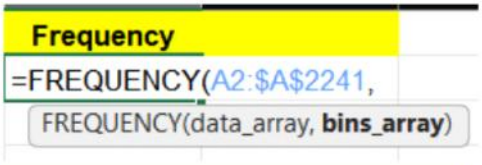
1. Visuals > Histogram.
2. Pivot Table -> Continuous Variable [Group] -> Histogram

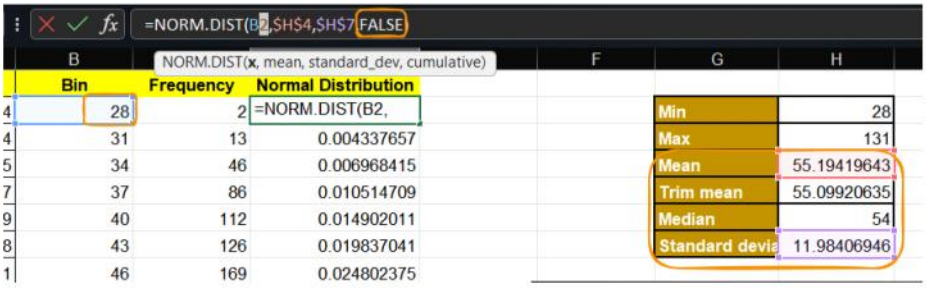
3. Manual Binning + Frequency Table



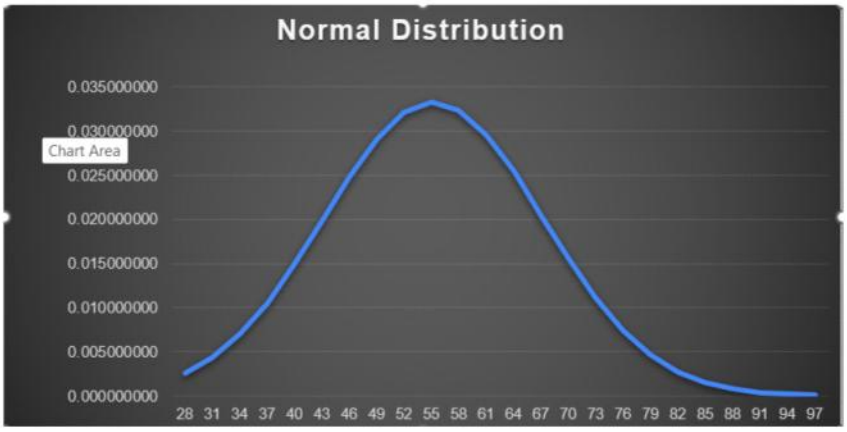
Min	28
Max	131
Mean	55.19419643
Trim mean	55.09920635
Median	54
Standard devia	11.98406946



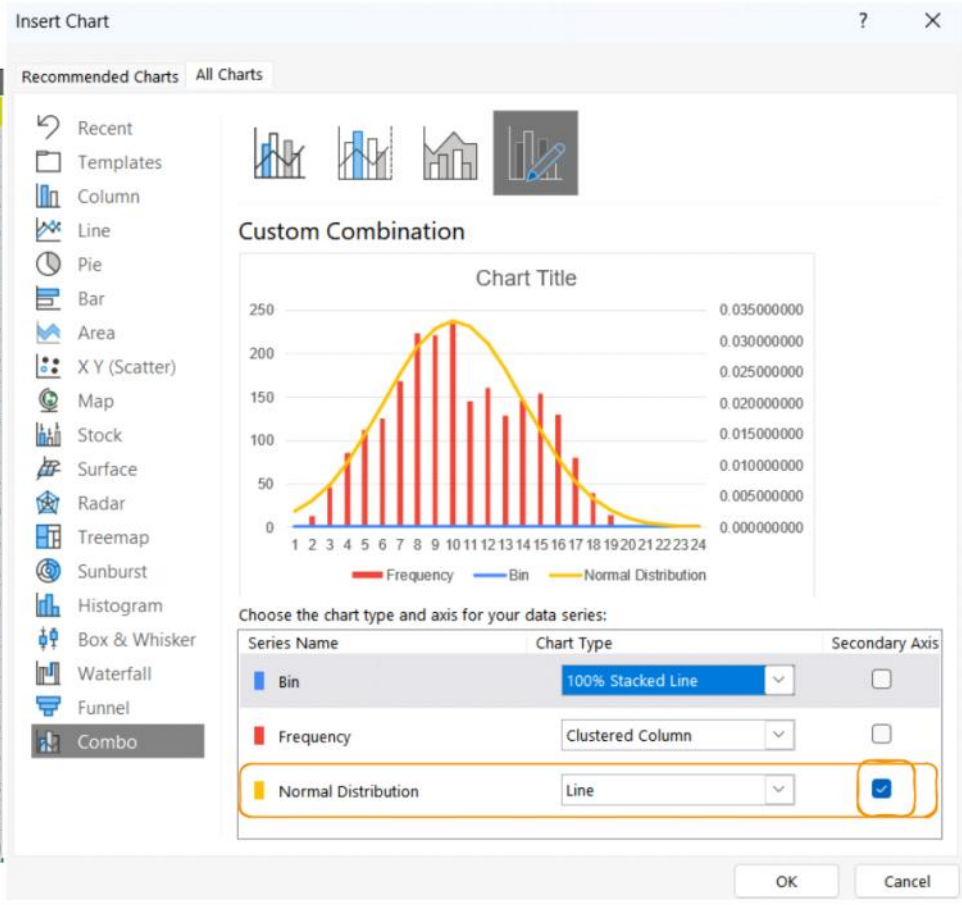


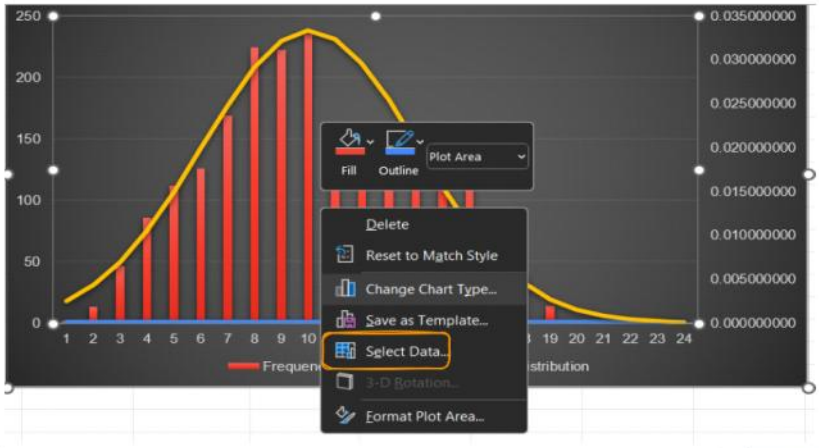


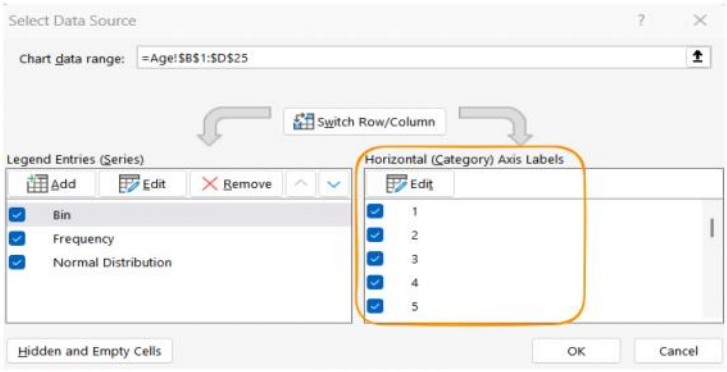
B	C
Bin	Normal Distribution
28	0.002536067
31	0.004337657
34	0.006968415
37	0.010514709
40	0.014902011
43	0.019837041
46	0.024802375
49	0.029126885
52	0.032127667
55	0.033285013
58	0.032389381
61	0.029603356
64	0.025413450
67	0.020491356
70	0.015518943
73	0.011039209
76	0.007375614
79	0.004628527
82	0.002728172
85	0.001510376
88	0.000785385
91	0.000383588
94	0.000175967
97	0.000075820



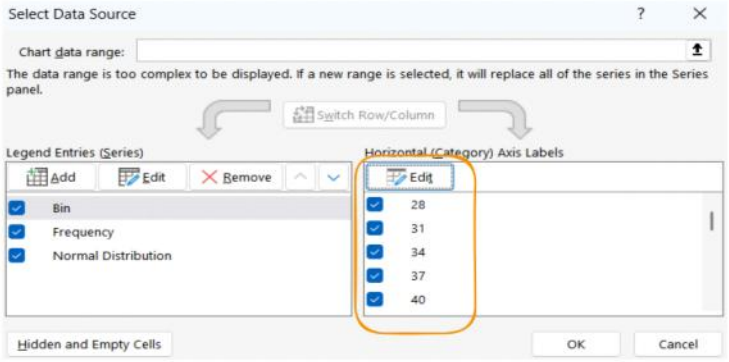
B	C	D
Bin	Frequency	Normal Distribution
28	2	0.002536067
31	13	0.004337657
34	46	0.006968415
37	86	0.010514709
40	112	0.014902011
43	126	0.019837041
46	169	0.024802375
49	224	0.029126885
52	222	0.032127667
55	235	0.033285013
58	145	0.032389381
61	161	0.029603356
64	129	0.025413450
67	147	0.020491356
70	154	0.015518943
73	130	0.011039209
76	80	0.007375614
79	40	0.004628527
82	14	0.002728172
85	2	0.001510376
88	0	0.000785385
91	0	0.000383588
94	0	0.000175967
97	0	0.000075820

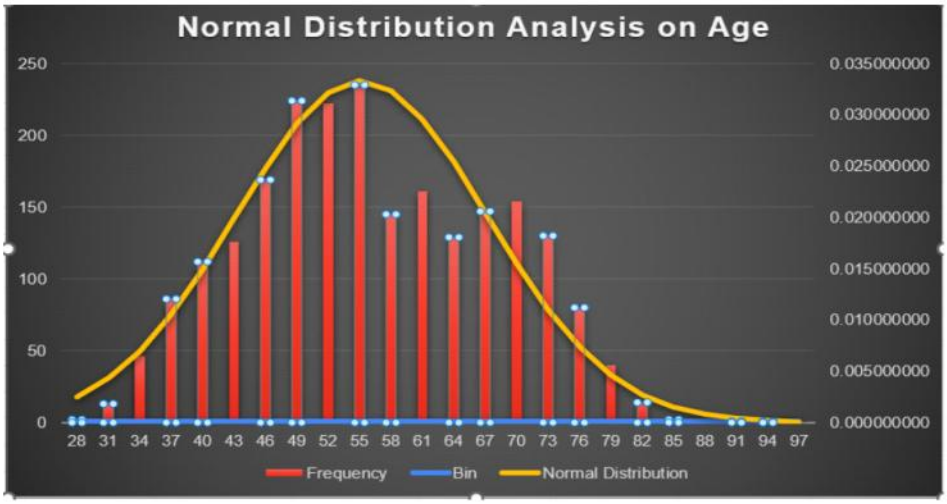



Fill
Outline
Plot Area
Delete
Reset to Match Style
Change Chart Type...
Save as Template...
Select Data...
3-D Rotation...
Format Plot Area...


Select Data Source
Chart data range: =Age!$B$1:$D$25
Switch Row/Column
Legend Entries (Series)
Add
Edit
Remove
Bin
Frequency
Normal Distribution
Horizontal (Category) Axis Labels
Edit
1
2
3
4
5
Hidden and Empty Cells
OK
Cancel


Axis Labels
Axis label range:
=Age!$B$2:$B$25
= 28, 31, 34, 37...
OK
Cancel


Select Data Source
Chart data range:
The data range is too complex to be displayed. If a new range is selected, it will replace all of the series in the Series panel.
Switch Row/Column
Legend Entries (Series)
Add
Edit
Remove
Bin
Frequency
Normal Distribution
Horizontal (Category) Axis Labels
Edit
28
31
34
37
40
Hidden and Empty Cells
OK
Cancel


Normal Distribution Analysis on Age
Frequency
Bin
Normal Distribution

How do we draw conclusions from seeing all three visuals in one place?

- Most of the audience are from middle age between 30 - 70.
- The Age column being done with univariate analysis shows symmetric distribution.

Univariate - One Variable
- Categorical - Pivot Table
- Continuous - Histogram / Frequency
  - TRIMMEAN [Outlier Detection]

Bivariate Analysis - Two Variables
[One Dependent Variable, Independent Variable]

Correlation	Strength
0.00–0.19	Very Weak
0.20–0.39	Weak
0.40–0.59	Moderate
0.60–0.79	Strong
0.80–1.00	Very Strong

Study Time(hrs)	Exam Score(%)
3	75
5	90
1	50
6	95
4	85
7	98

CORRELATION	0.964573	96%

INDEPENDENT VARIABLE

DEPENDENT VARIABLE

Hours of TV Watch	Number of Books Read
10	5
20	4
8	6
12	3
16	2
11	4

CORRELATION	-0.57844	-58%

Bivariate Analysis → Category , Continuous

2 2

category, continuous
category, category
continuous, category
continuous, continuous

→ Categorical , Categorical

Categorical , Continuous

Continuous , Continuous

Categorical , Categorical

Pivot Table -> Row, Column [@d Matrix]

- Age Bracket VS Bike Purchased.
- Marital Status VS Region
- Education Vs Occupation

Count of ID	Column Labels					
Row Labels	Clerical	Management	Manual	Professional	Skilled Ma	Grand Total
Bachelors	4.90%	9.70%	0.20%	9.20%	6.60%	30.60%
Graduate Degre	2.40%	5.90%	0.60%	4.50%	4.00%	17.40%
High School	0.40%	1.20%	4.40%	5.50%	6.40%	17.90%
Partial College	7.60%	0.50%	3.60%	8.00%	6.80%	26.50%
Partial High Sch	2.40%	0.00%	3.10%	0.40%	1.70%	7.60%
Grand Total	**17.70%**	**17.30%**	**11.90%**	**27.60%**	**25.50%**	**100.00%**

Count of ID	Column Labels		
Row Labels	Female	Male	Grand Total
Married	23.90%	29.90%	53.80%
Single	25.00%	21.20%	46.20%
Grand Total	**48.90%**	**51.10%**	**######**

Count of ID	Column Labels			
Row Labels	Europe	North America	Pacific	Grand Total
Married	14.60%	29.70%	9.50%	53.80%
Single	15.40%	21.10%	9.70%	46.20%
Grand Total	**30.00%**	**50.80%**	**19.20%**	**100.00%**

Categorical , Continuous

Education Level VS Income

Gender VS Age

Region VS Total Spent

Martial Status VS Total Purchases

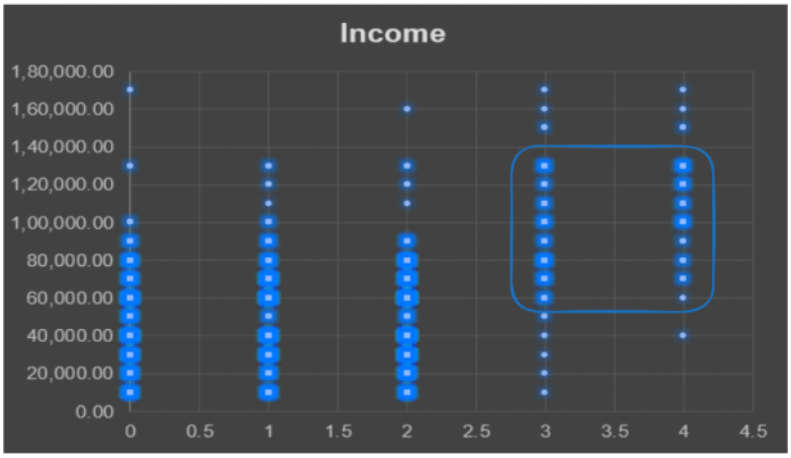
Row Labels	Average of Income
Clerical	31073
Management	86647
Manual	16723
Professional	75072
Skilled Manual	51608
Grand Total	**56360**

Row Labels	Sum of TotalMnt
2n Cycle	12831
Basic	1254
Graduation	70034
Master	15847
PhD	19572
Grand Total	**119538**

Continuous , Continuous

Correlation [Make a habit to check corr]

	P	Q	R
	Cars	Income	Correlation
0		40,000.00	43%
1		30,000.00	
2		80,000.00	
1		70,000.00	
0		30,000.00	
0		10,000.00	
4		1,60,000.00	
0		40,000.00	
2		20,000.00	
1		1,20,000.00	
2		30,000.00	
4		90,000.00	
0		1,70,000.00	
1		40,000.00	
1		60,000.00	
1		10,000.00	
2		30,000.00	
0		30,000.00	
1		40,000.00	
2		20,000.00	
0		40,000.00	
4		80,000.00	
0		40,000.00	
3		80,000.00	



Row Labels	Average of Income
0	47530.36437
1	50823.97004
2	49298.24561
3	91764.70588
4	108305.0847
Grand Total	**56360**

P	Q	R
Total kids	TotalMnt	Correlation
2	11	-44%
0	129	
0	73	
1	106	
1	31	
1	7	
0	170	
1	291	
0	3	
1	93	
1	149	
1	3	
1	15	
1	15	
1	16	
1	21	
0	6	
0	19	
2	8	
1	29	
2	5	
2	2	
0	160	
2	4	

"The more kids they have, less amount they spent"

T	U	V
Income	Total purchases	Correlation
55158	12	53%
52203	19	
82576	14	
73803	15	
7500	5	
7500	6	
32632	12	
75484	13	
27469	3	
53230	17	
64176	16	
49431	12	
62972	10	
62972	10	
56937	17	
27683	10	
22304	3	

Moderate Correlation -> "Higher the income, more chance to spend money on both store."

Row Labels	Sum of NumWebPurchase	Sum of NumStorePurchases
28-37	493	857
38-47	1661	2355
48-57	2897	4010
58-67	2104	2974
68-77	1745	2464
78-87	243	302
118-127	6	6
128-137	1	2
Grand Total	**9150**	**12970**

More people within the middle age group are looking to buy the products from stores.

Conclude the Data Analysis

1. Importance of Effective Communication.
   - 1.1 - Understanding the stakeholder need.
   - 1.2 Driving the Decision Making
   - 1.3 Ensuring the actionable insights
   - 1.4 Adapting the stakeholder preferences.
   - 1.5 Avoiding Misinterpretation.

2. Techniques for Communicating insights
   - 2.1 Simplify the complex Information.
   - 2.2 Use Visualization.
   - 2.3 Tell A Story
   - 2.4 Provide the context
   - 2.5 Highlight the key findings.
   - 2.6 Use Examples And Analogy.
   - 2.7 Use clear and concise Language

codingninjas

Analysing the customer behaviour for ABC Company Ltd.

Identifying target for new product launch

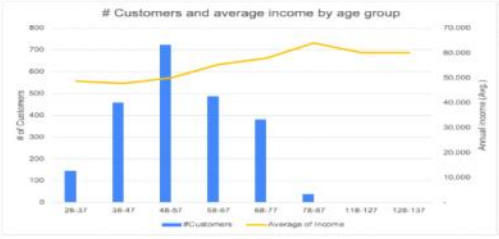
Analysis by XYZ | 31st March, 2024

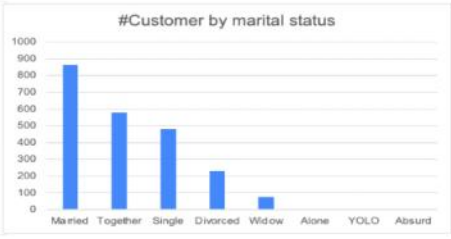
Understanding the distribution of customers currently

Distribution of customers and income by age, marital status and education

Key Observations

- Most of the customers are middle aged ranging from 35 to 60 years, shows potential of targeting this group often
- Married people are the highest customers, this might be because of high number of kids
- Graduated people are ideal customers maybe because of higher income and more kids





Average income and # customers by age group

48-57 age, married and graduated people majorly become customers of ABC company Ltd.

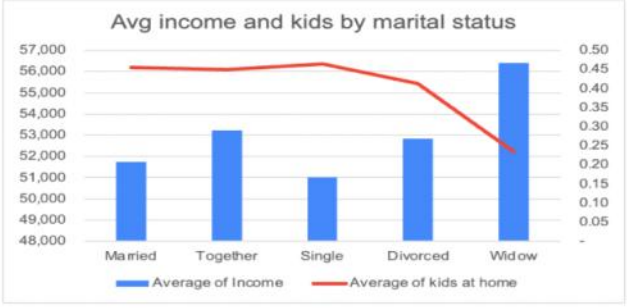
Data from 2012 to 2014 | Data source: ABC Company Ltd. | Verified customers of ABC included in the analysis

We will evaluate reason why customer becomes a customer

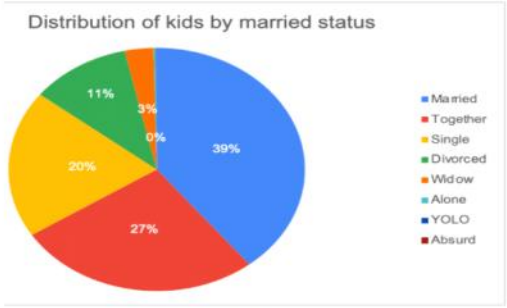
We will explore if income and count of kids impact on married behaviour

Key Observations

- Married people have higher kids at home which means kids at home result in customer conversion
- Higher distribution (40%) of kids for married couple means kids drive purchases for these customers



Average kids at home and income by marital status



High distribution of kids for married people

Average number of kids have imp… …come doesn't impact customer

Data from 2012 to 2014 | Data source: ABC Company Ltd. | Verified customers of ABC included in the analysis

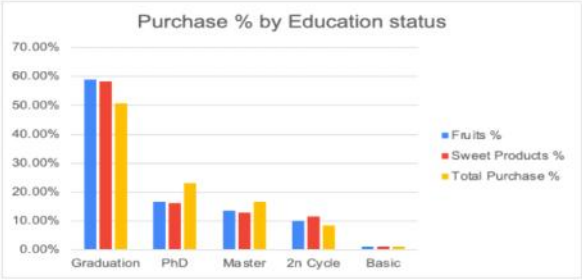
We will evaluate if high customer result in high purchase

Married and graduate were highest customer segment, assessing purchases here

Key Observations

- Married people are the highest customer count and also result in highest purchase (±40%)
- Similar to married people, graduates also were highest customer count and result in high purchases

Distribution of purchases by marital status



Distribution of purchase by education

High customer count also result in higher purchase, which means customer count is correlated with purchase

codingninjas

New Product Launch

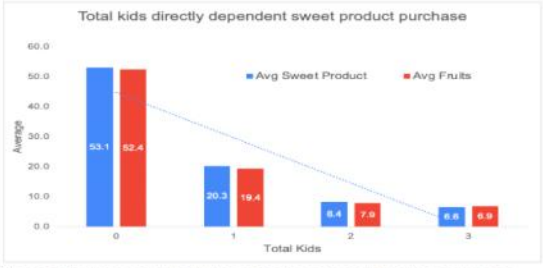
Identifying best target customer group

Analysing best customer segment for new sweet product

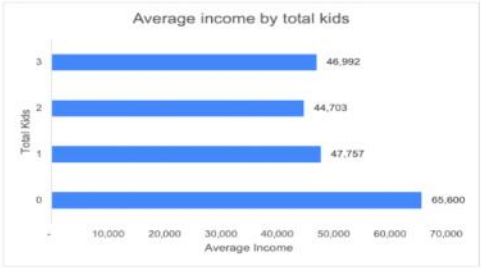
Kids at home and annual income are correlated and result in higher sweet purchas

Key Observations

- We analyzed multiple factors to identify relationship between sweet products and sales
- Eventually, we identified that sales of both sweet products and fruit products are dependent on total kids
- Average income is also a factor of total kids



0 kids at home result in highest average sweet product purchase



Lesser kids at home means higher average income

For new product launch, we should target people with 0 total kids and also groups with higher income

Data from 2012 to 2014 | Data source: ABC Company Ltd. | Verified customers of ABC included in the analysis

Summary

Based on our **observations**, we can conclude the following for ABC Company Ltd. -

- People with graduation as their education status are more prone to become customers
- Married customers with kids should also be an area of focus for the company
- Average number of kids impact customer behaviour positively
- Average income does not impact the customer conversion
- Number of customers are directly related to the purchases made

- For **new product launch,** we should prioritise targeting people with "0" **kids at home** and people with **higher average income**

Communication Template

codingninjas

Introduction to the business problem and objective

Additional details of the problem (if any)

Analysis Owner | Date

High-level summary of the problem and findings

- What was the problem
- How did we approach the problem
- What are the high level findings about the problem
- Recommendations - Next Steps

Talking Header to talk about the business objective / findings

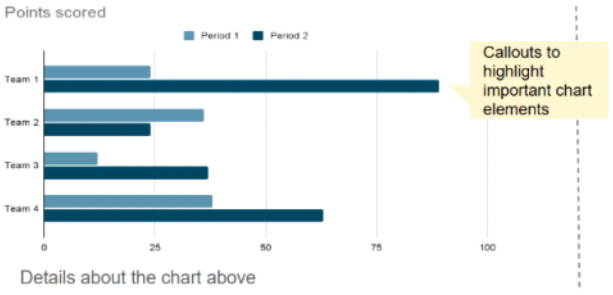
Sub- Header provides additional details about the objective, solution

Insights Header (eg. "Key Observations")

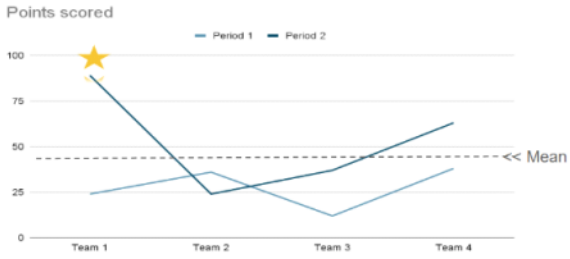
- Important Insights 1
- Insights 2
- Insights 3
- Insights 4

Insights Header

- Key points to consider
- Key mappings
- Other trends
- etc.



Details about the chart above



Details about the chart above

Takeaway Box concludes the entire slide with actionable insights and recommendations

Key highlight about data | Data time | Data source | Data caveats | …

Key information about the analysis | Mappings | Assumptions | etc..