**A Data Analytics Report Submitted to**



**Rajiv Gandhi Proudyogiki Vishwavidyalaya, Bhopal Towards Partial Fulfillment for the Award of**

**Bachelor of Technology (Computer Science and Engineering)**

**Under the Supervision of Prof. Anurag Punde Sir**

**Submitted By Balraj Singh Ranawat**



**Department of Computer Science and Engineering Acropolis Institute of Technology & Research, Indore Jan-June2024**

|  |  |  |
| --- | --- | --- |
| **S.No.** | **Experiment** | **Remarks** |
| 1. | Data Analysis Questions:   1. Data Analysis Principles 2. Statistical Analytics 3. Hypothesis Testing 4. Regression 5. Correlation 6. ANOVA |  |
| 2. | Dashboards:   1. Exploring Car Dataset 2. Cookie Data: Trends and Analysis Report 3. Exploring Loan Dataset 4. Exploring sales on different states of US 5. Store Data Analysis 6. Shop Sale Data Report 7. Sale Samples: A Detailed Report |  |
| 3. | Reports:   1. Exploring Car Dataset 2. Cookie Data: Trends and Analysis Report 3. Exploring Loan Dataset 4. Exploring sales on different states of US 5. Store Data Analysis 6. Shop Sale Data Report 7. Sale Samples: A Detailed Report |  |
| 4. | Remote Ratio Forecast Analysis (2020-2026) |  |

Introduction

## Car Collection Data Report

### The Car Collection dataset provides a detailed examination of various attributes of different car models, including make, model, color, mileage, price, and cost. This report aims to analyze and extract insights from the dataset to support decision-making processes related to car purchases and understanding market trends. The dataset includes a total of six cars with different models: Honda, Chevrolet, Nissan, Toyota, Dodge, and Ford.

### The primary audience for this report includes car enthusiasts, automotive industry professionals, analysts, and individuals interested in exploring car market trends. The scope of this report includes a thorough analysis of the dataset, featuring statistical evaluations, visualizations, and interpretation of results.

### Throughout the analysis, we have addressed several key questions and conducted corresponding analyses to uncover meaningful insights

### .Questionnaire

1. Compare the mileage of Chevrolet Impala to Toyota Corolla. Which of the two is giving best mileage?
2. Justify, Buying of any Ford car is better than Honda.
3. Among all the cars which car color is the most popular and is least popular?
4. Compare all the cars which are of silver color to the green color in terms of Mileage.
5. Find out all the cars, and their total cost which is more than $2000?

### Analytics

1. Compare the mileage of Chevrolet Impala to Toyota Corolla. Which of the two is giving best mileage?

This analysis compares the fuel efficiency (mileage) of two popular car models, Chevrolet Impala and Toyota Corolla. For performing this the dataset was filtered to isolate data and column chart was created And based on the analysis it was concluded that Chevrolet Impala(114243) provides better mileage compared to Toyota Corolla(92377).

|  |  |  |
| --- | --- | --- |
| Average of Mileage | |  |
| Make | Model | Total |
| Chevrolet | Impala | 114243 |
| Chevrolet Total | | 114243 |
| Toyota | Corolla | 92377 |
| Toyota Total | | 92377 |

120000

Compare the mileage of Chevrolet Impala to Toyota Corolla

114243

100000

92377

80000

60000

40000

20000

0

Impala

Corolla

Chevrolet

Toyota

101123.

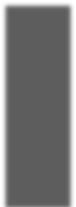
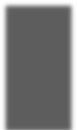
4

Grand Total

1. Justify, Buying of any Ford car is better than Honda.

This analysis aims to provide justification for purchasing any Ford car over Honda by comparing their respective attributes, specifically focusing on price considerations.

But, after the analysis performed on the dataset it was not justifying the statement rather the Honda cars have better average mileage(89623.3) and average price(3193.6) as compared to Ford cars.



**Buying any Ford car is better than Honda**

100000

90000

80000

70000

60000

50000

40000

30000

20000

89623.2

55452

Average of Price

Average of Mileage

10000

0

3581.6

3193.6

Ford Honda

1. Among all the cars which car color is the most popular and is least popular?

This analysis aims to identify the most popular and least popular car colors among all the cars in the dataset based on the count of the make.

The analysis showed that the most popular color of the cars are Black and White both having the 25% each of the making by company whereas the Green and Blue cars are at the least popular cars with both having the 12% of making.



**POPULAR COLOR CAR AMONG ALL THE CARS**

**White, 3, 13%**

**Black, 6, 25%**

**Silver, 6, 25%**

**Blue, 3, 12%**

**Red, 3, 13%**

**Green, 3, 12%**

1. Compare all the cars which are of silver color to the green color in terms of Mileage.

This analysis aims to identify the cars which are of silver colour to the green colour in terms of mileage and the insights are there are 5 silver cars namely, Mustang, Impala, Corolla, Charger, Accord where Accord have the highest Average mileage among them which is 101354. And 2 green cars namely, Silverado and Altima where Silverado had the highest mileage of 109231.

**Comparison of all the cars which are silver-colored to green-**

**colored in terms of Mileage**

Mustang

41560

Impala

87675

Corolla

59169

Charger

46513

Accord

101354

Silverado

109231

Altima

62540

0

20000

40000

60000

80000

100000

120000

Silver

1. Find out all the cars, and their total cost which is more than $2000?

Green

This analysis aims to find out the car’s costs more than $2000. And by using bar graph and taking value as sum of cost it shows the desired result. The total grand cost of all cars exceeding

$2000 is $66150.

Ford

Honda Nissan Toyota

### Conclusion and Review



**Total cost of cars exceeding $2000**

Corolla

Camry Maxima Altima CRV

Civic Accord Mustang Fusion

F-150

Escape Charger Silverado Malibu

Impala

0

1000 2000 3000 4000 5000 6000 7000 8000 9000 10000

Do

dg Chevrolet e

Comparison: The analysis comparing the mileage of Chevrolet Impala and Toyota Corolla revealed that Chevrolet Impala provides better fuel efficiency.

Ford vs. Honda Comparison: Contrary to the initial assumption, the analysis did not support the claim that Ford cars are better than Honda cars in terms of mileage and price. Honda cars were found to have better average mileage and price compared to Ford cars.

Popular Car Colors: The analysis identified Black and White as the most popular car colors, each comprising 25% of the car production. Conversely, Green and Blue were found to be the least popular colors, each accounting for only 12% of car production.

Silver vs. Green Cars Comparison: Among silver-colored cars, Accord exhibited the highest average mileage, while Silverado had the highest mileage among green-colored cars.

Cars Costing more than $2000: The analysis determined that the total cost of cars exceeding

$2000 amounted to $66150.

The analysis provided valuable insights into various aspects of the dataset, including mileage comparisons, car color popularity, and cost considerations. However, there were discrepancies between the initial assumptions and the findings, particularly in the comparison between Ford and Honda cars. The analysis was thorough and utilized appropriate visualizations, such as column charts and bar graphs, to present the findings effectively. Overall, the report offers valuable information for car buyers, industry professionals, and researchers interested in understanding trends within the car market. However, it's important to note the limitations of the analysis, such as the dataset's completeness and the need for further exploration into other factors influencing car purchasing decisions.

### Regression

The Regression Analysis table provides insights into the relationship between the dependent variable (Mileage) and the independent variables (Cost and Price) in the Car Collection Dataset. The analysis indicates a strong positive correlation (Multiple R = 0.962639) between

the variables, with approximately 92.67% of the variance in Mileage explained by Cost and Price (R Square = 0.926673). Both Cost and Price have statistically significant effects on Mileage, as evidenced by their coefficients and low p-values. Specifically, Price demonstrates a stronger impact on Mileage compared to Cost. The ANOVA table further confirms the overall significance of the regression model, with a small p-value (1.22E-12) for the F-statistic. Overall, the regression analysis suggests that both Cost and Price play significant roles in determining Mileage in the car collection dataset.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *Regression Statistics* | |  | | | | | | |
| Multiple R | 0.962639 |
| R Square | 0.926673 |
| Adjusted R Square | 0.91969 |
| Standard Error | 259.2716 |
| Observation  s | 24 |
| ANOVA | | | | | | | | |
|  | *df* | *SS* | *MS* | *F* | *Significanc*  *e F* |  |  |  |
| Regression | 2 | 1783989  7 | 891994  8 | 132.694  3 | 1.22E-12 |  |  |  |
| Residual | 21 | 1411657 | 67221.7  8 |  |  |  |  |  |
| Total | 23 | 1925155  4 |  |  |  |  |  |  |
|  | | | | | | | | |
|  | *Coefficient s* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* | *Lower 95.0%* | *Upper 95.0%* |
| Intercept | 441.3528 | 288.784  8 | 1.52831 | 0.14135  9 | -159.208 | 1041.91  4 | - 159.208 | 1041.91  4 |
| X Variable 1 | -0.00058 | 0.00169  9 | -  0.34395 | 0.73430  4 | -0.00412 | 0.00294  9 | -  0.00412 | 0.00294  9 |
| X Variable 2 | 1.038413 | 0.07049  2 | 14.7308  4 | 1.52E-  12 | 0.891816 | 1.18501 | 0.89181  6 | 1.18501 |

### Anova: one factor

In the Single Factor ANOVA table, there are three groups: Mileage, Cost, and Price. For the Mileage group, there are 24 observations, with a sum of 2011267, an average of 83802.79, and a variance of 1.21E+09. Similarly, for the Cost group, there are 24 observations, with a sum of 66150, an average of 2756.25, and a variance of 705502.7. For the Price group, there are also 24 observations, with a sum of 78108, an average of 3254.5, and a variance of 837024.1.

The ANOVA section assesses the differences in means among the levels of the single factor variable (Mileage, Cost, and Price). The Between Groups SS (Sum of Squares) is 1.04E+11 with 2 degrees of freedom (df) and a Mean Squares (MS) of 5.22E+10. This indicates significant variation among the group means. The F-statistic is 128.8822, with a p-value of 5E- 24, which is much lower than the typical significance level of 0.05, suggesting a highly significant difference in means among the groups. The Within Groups SS is 2.8E+10 with 69 df. The Total SS is 1.32E+11.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Anova: Single Factor | | | | |  | |
|  | | | | |
| SUMMARY | | | | |
| *Groups* | *Count* | *Sum* | *Average* | *Variance* |
| Mileage | 24 | 2011267 | 83802.79 | 1.21E+09 |
| Cost | 24 | 66150 | 2756.25 | 705502.7 |
| Price | 24 | 78108 | 3254.5 | 837024.1 |
| ANOVA | | | | | | |
| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Between Groups | 1.04E+11 | 2 | 5.22E+10 | 128.8822 | 5E-24 | 3.129644 |
| Within Groups | 2.8E+10 | 69 | 4.05E+08 |  |  |  |
|  | | | | | | |
| Total | 1.32E+11 | 71 |  |  |  |  |

### Anova: two factor

In the Two Factor ANOVA table, there are two factors: Rows and Columns, with respective levels of Mileage, Cost, and Price. The summary section provides the count, sum, average, and variance for each row and column combination. For example, Row 1 represents Mileage with a count of 3, a sum of 70512, an average of 23504, and a variance of 1.2E+09. Similarly, Row 2 represents Cost with a count of 3, a sum of 99635, an average of 33211.67, and a variance of 2.88E+09.

The ANOVA section assesses the sources of variation in the data. The Rows SS (Sum of Squares) is 8.95E+09 with 23 degrees of freedom (df) and a Mean Squares (MS) of 3.89E+08. The Columns SS is 1.04E+11 with 2 df and an MS of 5.22E+10. Both Rows and Columns have p-values greater than 0.05, indicating that neither Rows nor Columns have a significant effect on the observed variances. The Error SS is 1.9E+10 with 46 df. The Total SS is 1.32E+11.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| *SUMMARY* | *Count* | *Sum* | *Average* | *Variance* |  | |
| Row 1 | 3 | 70512 | 23504 | 1.2E+09 |
| Row 2 | 3 | 99635 | 33211.67 | 2.88E+09 |
| Row 3 | 3 | 104854 | 34951.33 | 3.31E+09 |
| Row 4 | 3 | 79104 | 26368 | 1.77E+09 |
| Row 5 | 3 | 76673 | 25557.67 | 1.47E+09 |
| Row 21 | 3 | 47301 | 15767 | 5.38E+08 |
| Row 22 | 3 | 42702 | 14234 | 3.19E+08 |
| Row 23 | 3 | 66425 | 22141.67 | 9.74E+08 |
| Row 24 | 3 | 140665 | 46888.33 | 6.06E+09 |
|  | | | | |
| Mileage | 24 | 2011267 | 83802.79 | 1.21E+09 |
| Cost | 24 | 66150 | 2756.25 | 705502.7 |
| Price | 24 | 78108 | 3254.5 | 837024.1 |
| ANOVA | | | | | | |
| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Rows | 8.95E+09 | 23 | 3.89E+08 | 0.941208 | 0.549982 | 1.766805 |
| Columns | 1.04E+11 | 2 | 5.22E+10 | 126.3564 | 2.05E-19 | 3.199582 |
| Error | 1.9E+10 | 46 | 4.13E+08 |  |  |  |
| Total | 1.32E+11 | 71 |  |  |  |  |

### Descriptive Statistics

The descriptive statistics provide key metrics for the variables Mileage, Cost, and Price. For Mileage, the mean is 83802.79 with a standard error of 7112.652. The median is 81142, and the standard deviation is 34844.74, indicating a moderate spread of data around the mean. The range is 105958, with the minimum mileage recorded at 34853 and the maximum at 140811. The skewness is positive (0.386522), indicating a slight right skew in the distribution, while the kurtosis (-1.09718) suggests a relatively flat distribution.

For Cost, the mean is 2756.25 with a standard error of 171.4525. The median is 2750, and the standard deviation is 839.9421. The range is 3000, with the minimum cost at 1500 and the maximum at 4500. The skewness is positive (0.473392), indicating a slight right skew, and the kurtosis (-0.81266) suggests a relatively flat distribution.

For Price, the mean is 3254.5 with a standard error of 186.7512. The median is 3083, and the standard deviation is 914.8902. The range is 2959, with the minimum price at 2000 and the maximum at 4959. The skewness is positive (0.272019), indicating a slight right skew, and the kurtosis (-1.20291) suggests a relatively flat distribution. Overall, these statistics provide insights into the central tendency, variability, and distribution shape of the data for each variable.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Mileage* |  | *Cost* |  | *Price* |  |
|  |  |  |  |  |  |
| Mean | 83802.79 | Mean | 2756.25 | Mean | 3254.5 |
| Standard Error | 7112.652 | Standard Error | 171.4525 | Standard Error | 186.7512 |
| Median | 81142 | Median | 2750 | Median | 3083 |
| Mode | #N/A | Mode | 3000 | Mode | #N/A |
| Standard  Deviation | 34844.74 | Standard  Deviation | 839.9421 | Standard  Deviation | 914.8902 |
| Sample Variance | 1.21E+09 | Sample Variance | 705502.7 | Sample Variance | 837024.1 |
| Kurtosis | -1.09718 | Kurtosis | -0.81266 | Kurtosis | -1.20291 |
| Skewness | 0.386522 | Skewness | 0.473392 | Skewness | 0.272019 |
| Range | 105958 | Range | 3000 | Range | 2959 |
| Minimum | 34853 | Minimum | 1500 | Minimum | 2000 |
| Maximum | 140811 | Maximum | 4500 | Maximum | 4959 |
| Sum | 2011267 | Sum | 66150 | Sum | 78108 |
| Count | 24 | Count | 24 | Count | 24 |
|  |  |  |  |  |  |

### Correlation

The correlation coefficient between Cost and Price is -0.41106, indicating a moderate negative linear relationship between the two variables. This suggests that as the cost increases, the price tends to decrease, and vice versa. However, the strength of this correlation is moderate, meaning that the relationship is not extremely strong.

|  |  |  |
| --- | --- | --- |
|  | *Cost* | *Price* |
| Cost | 1 |  |
| Price | -0.41106 | 1 |
|  |  |  |

Introduction

# Order Data Report

This report delves into a comprehensive dataset capturing sales transactions within the automotive industry, encompassing various attributes such as Order ID, Order Date, Ship Date, Customer Details, Product Information, and Sales Figures. The primary objective of this analysis is to glean actionable insights to inform decision-making processes and drive business growth within the automotive sector. By examining sales data across different US states, segments, categories, and sub-categories, this report aims to identify key trends, top- performing segments, and areas of potential growth. Insights derived from this analysis will be invaluable for automotive industry stakeholders, including sales managers, marketers, and executives, seeking to optimize sales strategies, enhance customer satisfaction, and maximize revenue.

A R I Z O N A

C A L I F O R N I A F L O R I D A I L L I N O I S

K E N T U C K Y M I C H I G A N N E W Y O R K

O H I O P E N N S Y L V A N I A

T E X A S V I R G I N I A

W A S H I N G T O N W I S C O N S I N A R I Z O N A

C A L I F O R N I A F L O R I D A I L L I N O I S

K E N T U C K Y M I C H I G A N N E W Y O R K

O H I O P E N N S Y L V A N I A

T E X A S V I R G I N I A

W A S H I N G T O N W I S C O N S I N A R I Z O N A

C A L I F O R N I A F L O R I D A I L L I N O I S

K E N T U C K Y M I C H I G A N N E W Y O R K

O H I O P E N N S Y L V A N I A

T E X A S V I R G I N I A

W A S H I N G T O N W I S C O N S I N

### Questionnaire

1. Compare all the US states in terms of Segment and Sales. Which Segment performed well in all the states?
2. Find out top performing category in all the states?
3. Which segment has the most sales in the US, California, Texas, and Washington?
4. Compare total and average sales for all different segments?
5. Compare the average sales of different categories and subcategory of all the states.

### Analytics

1. Compare all the US states in terms of Segment and Sales. Which Segment performed well in all the states?

After comparing all the states in terms of segment and sales , California(222419.05) emerged as the state with the highest number of sales. Consumer(1148060.531) segment performed well in all the states.

COMPARE SALES ACROSS DIFFERENT SEGMENTS IN

EACH STATE

C O N S U M E R

C O R P O R A T E

H O M E O F F I C E

16415.078

222419.05

32675.948

44252.611

20430.72

36576.371

174100.923

42628.544

66818.653

95360.73

35683.63

73866.52

14232.36

11736.322

144731.4105

21539.9835

15984.28

7812.47

23391.553

74828.351

23112.647

30975.437

53357.3038

27501.48

36333.43

11454.45

7121.257

79156.003

34220.6005

18999.626

8215.2

16168.15

57431.873

9389.159

18482.56

19854.4984

7451.61

25006.9

5486.62

1. Find out top performing category in all the states?

Office Supplies is the top performing category in all the states with total count of sales of 5909 followed by furniture(2078) and technology(1813).

top-performing category in all the states

7000

6000

5000

4000

5909

3000

2000

1000

0

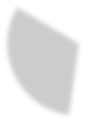
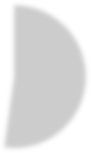
2078

1813

Furniture Office Supplies Technology

1. Which segment has most sales in US, California, Texas, and Washington?

Filtering the states for the total sales count and showing the percentage of distribution through pie chart. The consumer segment has the most sales in US, California, Texas, and Washington**.**



SOW THE DISTRIBUTION OF SALES AMONG DIFFERENT SEGMENTS IN US, CALIFORNIA, TEXAS, AND WASHINGTON.

**Home Office 17%**

**Consumer 52%**

**Corporate 31%**

1. Compare total and average sales for all different segments?

It is clearly visible that the consumer segment has higher average sales with 1148060.531 and home office segment has higher total sales of 243.40 followed by Corporate(233.1507195).



compare total and average sales for each segment

Home Office

243.4033086

424982.1769

Corporate

233.1507195

Average of Sales

688494.0748

Sum of Sales

Consumer

225.0657775

1148060.531

0

200000 400000 600000 800000 1000000 1200000 1400000

1. Compare average sales of different categories and subcategory of all the states.

The analysis shows the average sales for the 3 categories having multiple subcategories, the categories are Furniture, Office Supplies, Technology.

COMPARE THE AVERAGE SALES OF DIFFERENT

CATEGORIES AND SUBCATEGORIES.

F U R N I T U R E

O F F I C E S U P P L I E S

T E C H N O L O G Y

503.5982243

531.8331647

95.82386466

645.8937197

227.9268039

34.01963057

134.0675503

65.03244355

14.02785047

34.58746779

57.4202571

263.6338846

252.2842826

217.1781746

2215.880212

1645.553313

374.1808767

### Conclusion and Review

B O O K C A S E S

C H A I R S

F U R N I S H I N G S

T A B L E S

A P P L I A N C E S

A R T

B I N D E R S

E N V E L O P E S

F A S T E N E R S

L A B E L S

PAPER

S T O R A G E

S U P P L I E S

A C C E S S O R I E S

C O P I E R S

M A C H I N E S

P H O N E S

The analysis of sales data within the automotive industry reveals several key findings. California emerges as the top-performing state in terms of sales volume, with the Consumer segment demonstrating strong performance across all states. Office Supplies emerges as the top-performing category, followed by Furniture and Technology, indicating consumer preferences. The Consumer segment consistently dominates sales across the US, particularly in California, Texas, and Washington.

Additionally, the analysis highlights the higher average sales of the Consumer segment compared to the Home Office segment. Overall, these insights provide valuable guidance for optimizing sales strategies, improving customer engagement, and driving business success within the automotive industry.

### Regression

In this regression analysis for the Order dataset, there is almost no relationship between Order ID and Sales, as indicated by the very low multiple R and R-squared values (0.000434 and 1.88E-07, respectively). The coefficient for Order ID is not statistically significant, with a p- value of 0.965747. This suggests that Order ID does not predict Sales. Similarly, the ANOVA test confirms the lack of significance, with an F-statistic p-value of 0.965747.

|  |  |
| --- | --- |
| SUMMARY OUTPUT | |
|  | |
| *Regression Statistics* | |
| Multiple R | 0.000434 |
| R Square | 1.88E-07 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Adjusted R  Square | -0.0001 |  | | | | | | |
| Standard  Error | 625.334 |
| Observatio  ns | 9789 |
| ANOVA | | | | | |  | | |
|  | *df* | *SS* | *MS* | *F* | *Significan*  *ce F* |
| Regression | 1 | 721.16  37 | 721.16  37 | 0.0018  44 | 0.965747 |
| Residual | 9787 | 3.83E+  09 | 391042  .6 |  |  |
| Total | 9788 | 3.83E+  09 |  |  |  |
|  | | | | | | | | |
|  | *Coefficien*  *ts* | *Standar*  *d Error* | *t Stat* | *P-value* | *Lower*  *95%* | *Upper*  *95%* | *Lower*  *95.0%* | *Upper*  *95.0%* |
| Intercept | 230.5863 | 12.639  99 | 18.242  61 | 3.83E-  73 | 205.8093 | 255.36  33 | 205.80  93 | 255.36  33 |
| X Variable 1 | -9.6E-05 | 0.0022  35 | - 0.0429  4 | 0.9657  47 | -0.00448 | 0.0042  86 | - 0.0044  8 | 0.0042  86 |

### Descriptive Statistics

In the Sales dataset, the mean sales amount is 230.1162, with a standard error of 6.320053. The median sales value is 54.384, while the mode is 12.96. The standard deviation is 625.3021, indicating considerable variability in sales amounts. The data is highly positively skewed, with a skewness value of 13.05363, and exhibits high kurtosis at 307.3056, indicating heavy-tailed distribution. The range of sales values spans from 0.444 to 22638.48, with a total sum of 2252607 across 9789 observations.

|  |  |
| --- | --- |
| *Sales* | |
| Mean | 230.1162 |
| Standard Error | 6.320053 |
| Median | 54.384 |
| Mode | 12.96 |
| Standard Deviation | 625.3021 |
| Sample Variance | 391002.7 |
| Kurtosis | 307.3056 |
| Skewness | 13.05363 |
| Range | 22638.04 |
| Minimum | 0.444 |
| Maximum | 22638.48 |
| Sum | 2252607 |
| Count | 9789 |

Introduction

Cookie Data Report

In our cookie data set cookies—specifically six types: Chocolate Chip, Fortune Cookie, Sugar, oatmeal Raisin, Snickerdoodle, and White chocolate macadamia Nut. We've got a treasure trove of data on these cookies, covering how many units were sold, their costs, the money they brought in (revenue), and the profits they made. And we're not just looking at one place or time; we're exploring different countries and dates to see how things vary. This report isn't just about cookies; it's about understanding what people like, how much they're willing to pay, and where these treats are most popular. So, get ready to uncover some fascinating insights into the cookie world and what it means for businesses like yours.

Chocolate Chip

Fortune Cookie

25192

Oatmeal Raisin

Snickerdoodle

Sugar

White Chocolate

Macadamia Nut

Chocolate Chip

Fortune Cookie

Oatmeal Raisin

Snickerdoodle

Sugar

White Chocolate Macadamia Nut

Chocolate Chip

Fortune Cookie

Oatmeal Raisin

Snickerdoodle

Sugar

White Chocolate Macadamia Nut

### Questionnaire

1. Compare the profit earn by all cookie types in US, Malaysia, and India.
2. What is the average revenue generated by different types of cookies?
3. Which country sold most Fortune and sugar cookies in 2019 and in 2020?
4. Compare the performance of all the countries for the year 2019 to 2020. Which country perform in each of these years?
5. Which cookie category sold on the highest price, country wise and how much profit is earned by that category overall?

### Analytics

1. Compare the profit earn by all cookie types in US, Malaysia, and India.

This analysis compares the profit earned by all cookie types in US, Malaysia, and India. Max profit earned by India for chocolate chip followed by Malaysia and United States for the same.

**Compare the profit earned by each cookie type in the US,**

**Malaysia, and India.**

250000

200000

150000

100000

50000

0

India

Malaysia

United States

234576

81852.4

87012.5

72187.5

105511.25

189846

25403.2

79508.8

68060

45725.75

86255

208404

25290.4

99316

64547.5

60376.75

116223.25

1. What is the average revenue generated by different types of cookies?

This analysis aims to provide average revenue generated and it’s visible that white chocolate macadamia nut with average revenue generate is 8940.88 followed by chocolate chip.

Visualize the average revenue generated by each type of

cookie

White Chocolate Macadamia Nut 8940.880734

Sugar

4645.513761

Snickerdoodle

6316.086022

Oatmeal Raisin

8261.595745

Fortune Cookie

1658.075269

Chocolate Chip

8372.351485

0 1000 2000 3000 4000 5000 6000 7000 8000 9000 10000

1. Which country sold most Fortune and sugar cookies in 2019 and in 2020?

This analysis aims to compare the sales of fortune and sugar cookies in countries for 2019 and 2020, where India shows the significant sale in sugar cookies for the year 2020 having 30644 sales count, and the most sale sugar cookies in 2019 was from united kingdom followed by India then for the fortune cookie India again shows the higher sales of 25400 followed by Malaysia and for the fortune cookie Philippines shows the higher sales of 8782 followed by united states.

Compare the sales of Fortune and Sugar cookies in each

country for 2019 and 2020

29252

United States

5249

7961

23652

United Kingdom

11112

6525

24664

24758

22590

Philippines

8541

8782

19279

Sugar - 2020

Sugar - 2019

Fortune Cookie - 2020

Malaysia

6076

6922

20053

24832

Fortune Cookie - 2019

30644

India

10606

6090

25400

0

5000 10000 15000 20000 25000 30000 35000

1. Compare the performance of all the countries for the year 2019 to 2020. Which country perform in each of these years?

This analysis aims to compare the profit earned by countries in the financial year 2019 and 2020, according to the graph United kingdom shows the highest profit earned in 2020 with 471027.55 sales followed by United states with 456839.35 and the highest profit in 2019 was recorded by India with 155515.5 sales followed by Philippines with 131474.8.

Compare the profit earned by each country in 2019 and 2020

500000

450000

400000

350000

300000

250000

200000

150000

100000

50000

0

450816.15

471027.55

456839.35

373497.5

361529.75

2019

2020

India Malaysia Philippines United Kingdom United States

21255

Chocolate Chip

Fortune Cookie Oatmeal Raisin Snickerdoodle

Sugar

White Chocolate… Chocolate Chip Fortune Cookie Oatmeal Raisin Snickerdoodle

Sugar

White Chocolate… Chocolate Chip Fortune Cookie Oatmeal Raisin Snickerdoodle

Sugar

White Chocolate… Chocolate Chip Fortune Cookie Oatmeal Raisin Snickerdoodle

Sugar

White Chocolate… Chocolate Chip Fortune Cookie Oatmeal Raisin Snickerdoodle

Sugar

White Chocolate…

3803

25192

155515.5

17605

11408

9735

23106

14965

81852.4

87012.5

121301.25

2661

25403.2

13145

79508.8

11284 68060

870945725.75

17166

17565

29292448.8

21100

86255

131474.8

85722

11244 62360

8631 54479.25

124044.05

17796

19725

250215026.4

17110

88491.2

16976

85142.5

11625 62608

23988

22465

362275290.4

117318.55

16000

99316

13784 64547.5

8532

60376.75

17442

1. Which cookie category sold on the highest price, country wise and how much profit is earned by that category overall?

This analysis aims to find the cookie category sold for the highest price, country-wise, profit earned by that category, max of revenue is recorded by chocolate chip(23988) and sum of profit is recorded by sugar(2763364.45) for the country India followed by United Kingdom.

**Cookie category sold for the highest price, country-wise, profit**

**earned by that category overall.**

250000

200000

150000

100000

50000

0

India

Malaysia

Philippines

United Kingdom

United States

Max of Revenue Sum of Profit

234576

72187.5

105511.25

189846

167082

100912.5

214821

118982.5

208404

116223.25

### Conclusion and Review

The analysis provided insights into the profit earned by different cookie types in the US, Malaysia, and India. India emerged with the highest profit for chocolate chip cookies, followed by Malaysia and the United States.

White chocolate macadamia nut cookies generated the highest average revenue, followed closely by chocolate chip cookies.

In terms of sales, India showed significant sales of sugar cookies in 2020, while the United Kingdom had the highest sales of sugar cookies in 2019. For fortune cookies, India and Malaysia exhibited higher sales in both years, with the Philippines and the United States also contributing notable sales.

Regarding profit comparison by country for 2019 and 2020, the United Kingdom recorded the highest profit in 2020, followed by the United States. In 2019, India had the highest profit, followed by the Philippines.

Chocolate chip cookies were sold for the highest price in terms of revenue, while sugar cookies generated the highest profit overall.

The analysis presented valuable insights into the cookie industry, aiding stakeholders in understanding market dynamics and making informed decisions. The findings were effectively communicated through clear and appropriate visualizations. However, it's important to acknowledge the need for further exploration into additional factors influencing sales and profitability. Ensuring data accuracy and completeness is paramount for obtaining reliable insights.

### Regression

In the regression analysis for the Cookie dataset, the model's multiple R is 1, indicating a perfect linear relationship between the independent and dependent variables. The R-squared and adjusted R-squared values are both 1, indicating that the independent variables explain all the variability in the dependent variable. The standard error is very small (9.16E-12), suggesting precise estimates. The ANOVA results show that the regression model is highly significant (p < 0.05), with an F-statistic of 1.9E+31. The coefficients for the independent variables (X Variable 1, X Variable 2, X Variable 3) are all very close to 0, indicating no meaningful effect on the dependent variable. The p-values for these coefficients are all greater than 0.05, further supporting the lack of significance.

|  |  |
| --- | --- |
| SUMMARY OUTPUT | |
|  | |
| *Regression Statistics* | |
| Multiple R | 1 |
| R Square | 1 |
| Adjusted R  Square | 1 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Standard  Error | 9.16E-12 |  | | | | | | |
| Observation  s | 700 |
| ANOVA | | | | | |  | | |
|  | *df* | *SS* | *MS* | *F* | *Significanc*  *e F* |
| Regression | 3 | 4.78E+0  9 | 1.59E+0  9 | 1.9E+31 | 0 |
| Residual | 696 | 5.84E-20 | 8.39E-23 |  |  |
| Total | 699 | 4.78E+0  9 |  |  |  |
|  | | | | | | | | |
|  | *Coefficient s* | *Standar d Error* | *t Stat* | *P-value* | *Lower 95%* | *Uppe r 95%* | *Lowe r 95.0*  *%* | *Uppe r 95.0*  *%* |
| Intercept | -1.3E-11 | 7.3E-13 | -18.0657 | 4.09E-60 | -1.5E-11 | -1.2E-  11 | - 1.5E-  11 | -1.2E-  11 |
| X Variable 1 | 6.56E-17 | 8.42E-16 | 0.07789  2 | 0.93793  6 | -1.6E-15 | 1.72E  -15 | - 1.6E-  15 | 1.72E  -15 |
| X Variable 2 | 1 | 8.38E-16 | 1.19E+1  5 | 0 | 1 | 1 | 1 | 1 |
| X Variable 3 | -1 | 1.72E-15 | -5.8E+14 | 0 | -1 | -1 | -1 | -1 |

### Anova: one factor

The single-factor ANOVA analysis compares the variance between two groups: Cost and Profit. The Cost group comprises 700 observations, with a total sum of 1,926,955 and an average of 2,752.79. The Profit group also consists of 700 observations, with a total sum of 2,763,364 and an average of 3,947.66. The ANOVA results indicate that there is a significant difference between the means of the Cost and Profit groups (F = 90.92153, p < 0.05). This suggests that there is a statistically significant variation in the average values of Cost and Profit. The p-value (6.36E-21) is much smaller than the significance level (α = 0.05), indicating strong evidence against the null hypothesis. Therefore, we reject the null hypothesis and conclude that there is a significant difference in the mean values of Cost and Profit.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Anova: Single Factor | | | | |  | |
| SUMMARY | | | | |
| *Groups* | *Count* | *Sum* | *Average* | *Variance* |
| Cost | 700 | 1926955 | 2752.792 | 4149401 |
| Profit | 700 | 2763364 | 3947.664 | 6842519 |
| ANOVA | | | | |
| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Between Groups | 5E+08 | 1 | 5E+08 | 90.92153 | 6.36E-21 | 3.848119 |
| Within Groups | 7.68E+09 | 1398 | 5495960 |  |  |  |
| Total | 8.18E+09 | 1399 |  |  |  |  |

### Anova: two factor

The two-factor ANOVA without replication assesses the effects of two categorical independent variables, Revenue and Cost, on the dependent variable, Profit. The table provides a summary of the data for Revenue, Cost, and Profit, indicating the count, sum, average, and variance for each factor level. The ANOVA results reveal significant main effects for both Revenue (F = 14.75112, p < 0.05) and Cost (F = 1484.458, p < 0.05), as well as a significant interaction effect between Revenue and Cost (MS = 28507277, p < 0.05). The p-values for all factors are less than the significance level (α = 0.05), indicating strong evidence against the null hypothesis. Therefore, we reject the null hypothesis and conclude that both Revenue and Cost have a significant impact on Profit, and there is also a significant interaction effect between Revenue and Cost.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Anova: Two-Factor Without Replication | | | | | | | | |  | | |
|  | | | | | | | | |
| *SUMMARY* | *Count* | | *Sum* | | *Average* | | *Variance* | |
| Row 1 | 3 | | 17250 | | 5750 | | 6943125 | |
| Row 2 | 3 | | 21520 | | 7173.333 | | 10805909 | |
| Row 3 | 3 | | 23490 | | 7830 | | 12874869 | |
| Row 4 | 3 | | 12280 | | 4093.333 | | 3518629 | |
| Row 5 | 3 | | 13890 | | 4630 | | 4501749 | |
| Revenue | 700 | | 4690319 | | 6700.456 | | 21380458 | |
| Cost | 700 | | 1926955 | | 2752.792 | | 4149401 | |
| Profit | 700 | | 2763364 | | 3947.664 | | 6842519 | |
| ANOVA | | | | | | | | | | | |
| *Source of Variation* | | *SS* | | *df* | | *MS* | | *F* | | *P-value* | *F crit* |
| Rows | | 1.99E+10 | | 699 | | 28507277 | | 14.75112 | | 0 | 1.112595 |
| Columns | | 5.74E+09 | | 2 | | 2.87E+09 | | 1484.458 | | 0 | 3.002161 |
| Error | | 2.7E+09 | | 1398 | | 1932550 | |  | |  |  |
|  | |  | |  | |  | |  | |  |  |
| Total | | 2.84E+10 | | 2099 | |  | |  | |  |  |

### Descriptive Statistics

The descriptive statistics provide insights into the distribution and characteristics of the variables Unit Sold, Revenue, Cost, and Profit. For Unit Sold, the mean value is 1608.32 units, with a standard error of 32.79 units. The median value is 1542.5 units, indicating the central tendency of the data, and the mode is 727 units, representing the most frequently occurring value. The standard deviation, skewness, and kurtosis values indicate the dispersion, symmetry, and shape of the distribution, respectively. Similarly, for Revenue, Cost, and Profit, the descriptive statistics provide measures of central tendency, variability, and distributional characteristics. These statistics offer valuable insights into the distribution and variability of the variables, aiding in better understanding their underlying characteristics and informing further analysis.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| *Unit Sold* |  | *Revenue* |  | *Cost* |  | *Profit* |  |
|  |  |  |  |  |  |  |  |
| Mean | 1608.32 | Mean | 6700.456 | Mean | 2752.79  2 | Mean | 3947.66  4 |

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Standard  Error | 32.7865  2 | Standard  Error | 174.767 | Standard  Error | 76.9916  6 | Standard  Error | 98.8687  4 |
| Median | 1542.5 | Median | 5871.5 | Median | 2423.6 | Median | 3424.5 |
| Mode | 727 | Mode | 8715 | Mode | 3450 | Mode | 5229 |
| Standard  Deviation | 867.449  8 | Standard  Deviation | 4623.901 | Standard  Deviation | 2037.00  8 | Standard  Deviation | 2615.82  1 |
| Sample  Variance | 752469.  1 | Sample  Variance | 2138045  8 | Sample  Variance | 4149401 | Sample  Variance | 6842519 |
| Kurtosis | -0.31491 | Kurtosis | 0.464596 | Kurtosis | 0.81004  3 | Kurtosis | 0.33862  1 |
| Skewness | 0.43627 | Skewness | 0.867861 | Skewness | 0.93044  2 | Skewness | 0.84048  4 |
| Range | 4293 | Range | 23788 | Range | 10954.5 | Range | 13319 |
| Minimum | 200 | Minimum | 200 | Minimum | 40 | Minimum | 160 |
| Maximum | 4493 | Maximum | 23988 | Maximum | 10994.5 | Maximum | 13479 |
| Sum | 1125824 | Sum | 4690319 | Sum | 1926955 | Sum | 2763364 |
| Count | 700 | Count | 700 | Count | 700 | Count | 700 |

### Correlation

The correlation matrix reveals the relationships between the variables Unit Sold, Revenue, Cost, and Profit. A correlation coefficient close to 1 indicates a strong positive correlation, while a value near -1 signifies a strong negative correlation. For Unit Sold and Revenue, the correlation coefficient is approximately 0.796, indicating a moderately strong positive correlation. Similarly, Unit Sold and Profit exhibit a correlation coefficient of approximately 0.829, indicating a moderately strong positive relationship. Revenue and Cost demonstrate a correlation coefficient of around 0.992, signifying a strong positive correlation. Additionally, Revenue and Profit show a correlation coefficient of approximately 0.995, indicating a very strong positive relationship. Cost and Profit display a correlation coefficient of about 0.975, suggesting a strong positive correlation between these variables. These correlation values provide insights into the degree and direction of the relationships between the variables, aiding in understanding their associations and potential impacts on each other.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | *Unit Sold* | *Revenue* | *Cost* | *Profit* |
| Unit Sold | 1 |  |  |  |
| Revenue | 0.796298 | 1 |  |  |
| Cost | 0.742604 | 0.992011 | 1 |  |
| Profit | 0.829304 | 0.995163 | 0.974818 | 1 |

Introduction

## Loan Data Report

The loan dataset provides comprehensive information about loan applicants, encompassing attributes such as gender, marital status, education level, income details, loan amount, and property area. this dataset offers a rich source of insights into the dynamics of loan applications.

In this analysis, we aim to delve into the characteristics of loan applicants and explore patterns within the data. By leveraging pivot tables and charts, we seek to address specific queries regarding loan applicants' demographics, educational backgrounds, and loan amounts.

Understanding the nuances of loan applications is crucial for financial institutions to make informed decisions, optimize lending processes, and tailor services to meet the diverse needs of customers. Through this analysis, we endeavour to uncover actionable insights that can drive strategic decision-making and enhance the efficiency of loan management systems.

### Questionnaire

1. How many male graduates who are not married applied for Loan? What was the highest amount?
2. How many female graduates who are not married applied for Loan? What was the highest amount?
3. How many male non-graduates who are not married applied for Loan? What was the highest amount?
4. How many female graduates who are married applied for Loan? What was the highest amount?
5. How many male and female who are not married applied for Loan? Compare Urban, Semi- urban and rural based on amount.

### Analytics

1. How many male graduates who are not married applied for Loan? What was the highest amount?

This analysis shows the no. of male graduates applied for the loan and are not married with the highest amount. As of analysed the total no. of loan applied is 66 and max loan amount is 240.

Male graduates who are not married applied for a Loan and

the highest amount

300

250

240

200

150

Count of Loan\_ID

100

66

Max of LoanAmount

50

0

Total

1. How many female graduates who are not married applied for Loan? What was the highest amount?

This analysis shows the no. of female graduates applied for the loan and are not married with the highest amount. As of analysed the total no. of loan applied is 35 and max loan amount is 300.

Female graduates who are not married applied for a Loan

and the highest amount

400

300

300

200

Count of Loan\_ID

100

Max of LoanAmount

35

0

Total

1. How many male non-graduates who are not married applied for Loan? What was the highest amount?

This analysis shows the no. of male non-graduates applied for the loan and are not married with the highest amount. As the total no. of loan applied is 16 and max loan amount is 199.



Male non-graduates who are not married applied for

a Loan and the highest amount

199

Total

Max of LoanAmount

16 Count of Loan\_ID

0

50

100

150

200

250

1. How many female graduates who are married applied for Loan? What was the highest amount?

This analysis shows the no. of female graduates applied for the loan and are not married with the highest amount. As of analysed the total no. of loan applied is 21 and max loan amount is 460.



female graduates who are married applied for a Loan

and the highest amount

460

Total

21

Max of LoanAmount

Count of Loan\_ID

0

100

200

300

400

500

1. How many male and female who are not married applied for Loan? Compare Urban, Semi- urban and rural based on amount.

This analysis aims to compare the rural, semi urban, urban female and male who are not married and applied for the loan, where the no. is less in females but much higher in males

Females loan count in rural(1732),semiurban(1806), urban(1716) and males loan count in rural(3244),semiurban(3359),urban(3451).

males and females who are not married applied for a Loan,

Compare Urban, Semi-urban, and Rural on the basis of the loan amount

4000

3244

3359

3451

3000

2000

1732

1806

1716

1000

0

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Rural | Semiurban  Female | Urban | Rural | Semiurban  Male | Urban |

### Conclusion and Review

The analysis indicates clear gender disparities in loan applications. Male graduates not married dominated the applicant pool, followed by female graduates not married. Both male non- graduates not married and married female graduates also applied for loans, albeit in smaller numbers. Notably, males significantly outnumbered females across rural, semi-urban, and urban areas.

The analysis effectively illustrates gender-based trends in loan applications and provides valuable insights into borrower demographics. Further exploration into factors influencing loan decisions is recommended, along with visual enhancements to improve data presentation. Overall, the report lays a foundation for understanding loan dynamics, with potential for deeper insights.

### Regression

The regression analysis for the loan dataset reveals a multiple R coefficient of approximately 0.531, indicating a moderate positive relationship between the predictors and the loan amount. The R-squared value of around 0.282 suggests that about 28.2% of the variability in the loan amount can be explained by the independent variables. The coefficient for Applicant Income is approximately 0.096, and for Co-applicant Income, it's about 0.0068. These coefficients signify the impact of each predictor on the loan amount. Additionally, the ANOVA table displays a significant F-value of 37.32 (p < 0.05), affirming the statistical significance of the regression model. The ANOVA table indicates that the regression model is statistically significant, as evidenced by the low p-value (<0.05). This analysis provides insights into how applicant and co-applicant incomes influence the loan amount, aiding in better understanding the loan approval process.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| SUMMARY  OUTPUT |  | | | | | | | |
|  |
| *Regression Statistics* | |  | | | | | | |
| Multiple R | 0.531078  663 |
| R Square | 0.282044  546 |
| Adjusted R  Square | 0.274487  121 |
| Standard  Error | 50.85033  905 |
| Observatio  ns | 289 |
| ANOVA |  | | | | |  | | |
|  | *df* | *SS* | *MS* | *F* | *Significan*  *ce F* |
| Regression | 3 | 289502.8  035 | 96500.  93 | 37.320  19 | 2.25609E  -20 |
| Residual | 285 | 736940.7  397 | 2585.7  57 |  |  |
| Total | 288 | 1026443.  543 |  |  |  |
|  | | | | | | | | |
|  | *Coefficien*  *ts* | *Standard*  *Error* | *t Stat* | *P-*  *value* | *Lower*  *95%* | *Upper*  *95%* | *Lower*  *95.0%* | *Upper*  *95.0%* |
| Intercept | 66.69095  2 | 16.26833  015 | 4.0994  34 | 5.41E-  05 | 34.66963  005 | 98.71227  396 | 34.669  63 | 98.712  27 |
| X Variable  1 | 0.095771  273 | 0.045649  816 | 2.0979  55 | 0.0367  9 | 0.005917  708 | 0.185624  838 | 0.0059  18 | 0.1856  25 |
| X Variable  2 | 0.005807  787 | 0.000627  861 | 9.2501  22 | 5.49E-  18 | 0.004571  955 | 0.007043  619 | 0.0045  72 | 0.0070  44 |
| X Variable  3 | 0.006772  797 | 0.001264  765 | 5.3549  83 | 1.76E-  07 | 0.004283  331 | 0.009262  263 | 0.0042  83 | 0.0092  62 |

### Anova: one factor

In the ANOVA table for the single-factor analysis, the loan dataset is divided into two groups based on the factors Loan Amount and Loan Amount Term. The total sum of squares (SS) is approximately 8392703, with 2267909 within-group SS and 6124794 between-group SS. This results in a mean square (MS) of 3937.343 within groups and 6124794 between groups. The F-value of 1555.565 and the associated p-value of approximately 8.4E-166 indicate that there is a significant difference between the means of the two groups. Therefore, the factor being considered (Loan Amount vs. Loan Amount Term) has a significant impact on the loan dataset.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Anova: Single Factor | |  | | | | |
| SUMMARY | |
| *Groups* | *Count* | *Sum* | *Average* | *Variance* |  |  |
| Loan Amount | 289 | 39533 | 136.7924 | 3564.04 |  |  |
| Loan Amount Term | 289 | 99032 | 342.6713 | 4310.645 |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ANOVA |  |  |  |  |  |  |
| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Between Groups | 6124794 | 1 | 6124794 | 1555.565 | 8.4E-166 | 3.857654 |
| Within Groups | 2267909 | 576 | 3937.343 |  |  |  |
|  | | | | | | |
| Total | 8392703 | 577 |  |  |  |  |

### Anova: two factor

In the two-factor ANOVA without replication, the loan dataset is analyzed based on two factors: Loan Amount and Loan Amount Term. The dataset is arranged in rows and columns, with Loan Amount Term as the row factor and Loan Amount as the column factor. The total sum of squares (SS) is approximately 8392703, with 1264619 SS for rows, 6124794 SS for columns, and 1003290 SS for error. The mean square (MS) for rows is 4391.038, and for columns is 6124794. The F-value for rows is 1.260472, and for columns is 1758.156, with associated p-values indicating significance (p < 0.05). This suggests that both factors, Loan Amount and Loan Amount Term, have a significant impact on the loan dataset.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Anova: Two-Factor Without Replication | | | | | | |  | |
|  | | | | | | |
| *SUMMARY* | *Count* | *Sum* | *Average* | | *Variance* | |
| Row 1 | 2 | 470 | 235 | | 31250 | |
| Row 2 | 2 | 486 | 243 | | 27378 | |
| Row 3 | 2 | 568 | 284 | | 11552 | |
| Row 4 | 2 | 438 | 219 | | 39762 | |
| Row 5 | 2 | 512 | 256 | | 21632 | |
| Row 286 | 2 | 473 | 236.5 | | 30504.5 | |
| Row 287 | 2 | 475 | 237.5 | | 30012.5 | |
| Row 288 | 2 | 518 | 259 | | 20402 | |
| Row 289 | 2 | 278 | 139 | | 3362 | |
|  |  |  |  | |  | |
| Loan Amount | 289 | 39533 | 136.7924 | | 3564.04 | |
| Loan Amount Term | 289 | 99032 | 342.6713 | | 4310.645 | |
| ANOVA | | | | | | | | |
| *Source of Variation* | *SS* | *df* | *MS* | *F* | | *P-value* | | *F crit* |
| Rows | 1264619 | 288 | 4391.038 | 1.260472 | | 0.024978 | | 1.214301 |
| Columns | 6124794 | 1 | 6124794 | 1758.156 | | 1.2E-124 | | 3.87395 |
| Error | 1003290 | 288 | 3483.647 |  | |  | |  |
| Total | 8392703 | 577 |  |  | |  | |  |

### Descriptive Statistics

Descriptive statistics were computed for four variables in the loan dataset: Loan Amount Term, Applicant Income, Co-Applicant Income, and Loan Amount. For Loan Amount Term, the mean is approximately 342.67 months, with a standard error of 3.86 months. The median Loan Amount Term is 360 months, with a mode of 360 months as well. The standard deviation is

65.66 months, indicating variability in loan term lengths. Applicant Income has a mean of approximately 4637.35, with a standard error of 281.80. The median and mode are 3833 and 5000, respectively. The standard deviation is high at 4790.68, suggesting significant variability in applicant incomes. Co-Applicant Income has a mean of about 1528.26, with a standard error of 139.86. The median is 879, with a mode of 0, indicating a right-skewed distribution. The standard deviation is 2377.60, highlighting variability in co-applicant incomes. Lastly, Loan Amount has a mean of 136.79, with a standard error of 3.51. The median is 126, with a mode of 150. The standard deviation is 59.70, suggesting variability in loan amounts. These statistics provide insights into the central tendency, variability, and distribution of the loan dataset variables.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| *Loan Amount*  *Term* |  | *Applicant Income* |  | *Co- Applicant*  *Income* |  | *Loan Amount* |  |
| Mean | 342.671  3 | Mean | 4637.353 | Mean | 1528.263 | Mean | 136.7924 |
| Standard  Error | 3.86208  8 | Standard  Error | 281.8049 | Standard  Error | 139.8588 | Standard  Error | 3.51174 |
| Median | 360 | Median | 3833 | Median | 879 | Median | 126 |
| Mode | 360 | Mode | 5000 | Mode | 0 | Mode | 150 |
| Standard  Deviation | 65.6555 | Standard  Deviation | 4790.684 | Standard  Deviation | 2377.599 | Standard  Deviation | 59.69958 |
| Sample  Variance | 4310.64  5 | Sample  Variance | 22950653 | Sample  Variance | 5652978 | Sample  Variance | 3564.04 |
| Kurtosis | 8.62994 | Kurtosis | 141.612 | Kurtosis | 32.96701 | Kurtosis | 5.739804 |
| Skewness | -2.64147 | Skewness | 10.41123 | Skewness | 4.510775 | Skewness | 1.780616 |
| Range | 474 | Range | 72529 | Range | 24000 | Range | 432 |
| Minimum | 6 | Minimum | 0 | Minimum | 0 | Minimum | 28 |
| Maximum | 480 | Maximum | 72529 | Maximum | 24000 | Maximum | 460 |
| Sum | 99032 | Sum | 1340195 | Sum | 441668 | Sum | 39533 |
| Count | 289 | Count | 289 | Count | 289 | Count | 289 |

### Correlation

The correlation matrix for the loan dataset variables shows the correlation coefficients between Applicant Income, Co-Applicant Income, and Loan Amount. The correlation between Applicant Income and Co-Applicant Income is approximately -0.084, indicating a weak negative correlation between these two variables. The correlation between Applicant Income and Loan Amount is approximately 0.446, suggesting a moderate positive correlation. Similarly, the correlation between Co-Applicant Income and Loan Amount is approximately 0.230, indicating a weak positive correlation.

|  |  |  |  |
| --- | --- | --- | --- |
|  | *Applicant Income* | *Co-Applicant income* | *Loan Amount* |
| Column 1 | 1 |  |  |
| Column 2 | -0.08435 | 1 |  |
| Column 3 | 0.445695 | 0.230355 | 1 |

Introduction

# Shop Sales Data Report

This report delves into a comprehensive sales dataset, focusing on analysing sales performance and product trends among salesmen. The dataset comprises attributes such as salesmen details, product information, sales quantities, and profits earned. The primary objective of this analysis is to uncover insights that can inform sales strategy formulation and enhance business performance. By examining sales data over a specified period and comparing product performance, the report aims to identify top-performing salesmen, analyse product popularity, and understand sales trends. The insights derived from this analysis will be invaluable for sales managers, marketing professionals, and executives seeking to optimize sales strategies, maximize revenue, and drive business growth. Through this analysis, we aim to provide actionable insights that can guide decision-making and contribute to overall business success.

### Questionaries

1. Compare all the salesmen based on profit earn.
2. Find out most sold product over the period of May-September.
3. Find out which of the two product sold the most over the year Computer or Laptop?
4. Which item yield most average profit?
5. Find out average sales of all the products and compare them.

### Analytics

1. Compare all the salesmen on the basis of profit earn.

The comparison of all the salesmen on the basis of profit earned and the line chart shows that the rahul has the highest profit earned with value 493541.3255, compared to all the salesmen.



Compare all the salesmen on the basis of profit earned

500000

493541.3255

485039.1127

476120.3887

478167.1413

480000

460000

440000

420000

414776.4447

400000

380000

360000

Aman

Rahul

Ram

Rohit

Vinod

1. Find out most sold product over the period of May-September.

To identify the most sold product over the period of May-September, we would need to analyze the sales data within the timeframe. By aggregating the quantity sold for each product across all transactions during this period, and the most sold product over the period of May-September is Laptop with most sales in the September month with the value of 280.1970249.



SEP

AUG JUL JUN

MAY

Most sold product over the period of May-September.

**229.4219787**

**280.1970249**

**254.4439175**

**231.0465232**

**252.3313782**

**229.3404632**

**205.9430689**

**224.4657315**

**204.2370089**

**180.7583737**

**179.9459642**

**179.1335546**

**155.6549194**

**171.4153896**

**153.9488594**

0

50

100

150

200

250

300

Mobile Laptop

Computer

1. Find out which of the two product sold the most over the year Computer or Laptop?

The two product sold the most over the year between computer or laptop where Computer has the sold quantity of 2139.876313 and laptop has 2358.911786 units sold quantity.



COMPARE THE QUANTITY SOLD OF COMPUTERS AND LAPTOPS OVER THE YEAR

Laptop, 2358.911786

Computer, 2139.876313

4 . Which item yield most average profit?

This analysis shows that the Mobile has the most Average profit earned among Mobile, Laptop, and Computer where Mobile has the average profit earned of 7057.58477.

6600 6650 6700 6750 6800 6850 6900 6950 7000 7050 7100

|  |  |
| --- | --- |
|  | Compare the average profit earned from each item. |
| Mobile | 7057.58477 |
| Laptop | 6772.950369 |
| Computer | 6770.231898 |

5. Find out average sales of all the products and compare them.

The analysis shows that the average sales quantity of Laptop(19.49513873) is higher than the other products e.g. Mobile(19.41876737) and Computer(19.45342103).



**Compare the average sales quantity of each product.**

Mobile

19.41876737

Laptop

19.49513873

Computer

19.45342103

19.38

19.4

19.42

19.44

19.46

19.48

19.5

19.52

### Conclusion and Review:

The analysis reveals significant insights into sales performance and product trends among salesmen. Rahul emerges as the top performer, earning the highest profit compared to all other salesmen. Moreover, the most sold product over the period of May-September is identified as the Laptop, with the highest sales recorded in September. Between computers and laptops, laptops outperform computers in terms of units sold throughout the year. Additionally, mobile phones exhibit the highest average profit among mobiles, laptops, and computers. Lastly, laptops demonstrate the highest average sales quantity compared to mobiles and computers.

The analysis effectively highlights sales performance and product trends, providing valuable insights for sales strategy optimization. Visualizations aid in understanding trends over time and product popularity. However, deeper insights into factors influencing sales fluctuations and product preferences could enhance the analysis. Overall, the report offers actionable insights for improving sales strategies and maximizing revenue.

### Regression

This regression analysis demonstrates a robust relationship between the quantity of items sold (X Variable 1) and the corresponding sales amount (Y). With a high R Square value of approximately 0.91, it indicates that about 91% of the variability in sales amount can be accounted for by changes in the quantity of items sold. For each additional unit increase in the quantity sold, there's an average increase of approximately $246.47 in sales amount. Both the intercept and the coefficient of the X Variable 1 are statistically significant, with t-stats of 23.38 and 58.73, respectively, and very low p-values (nearly zero), affirming the reliability of these coefficients. Thus, we conclude that the quantity of items sold serves as a strong predictor of sales amount in this dataset.

|  |
| --- |
| SUMMARY OUTPUT |
|  |
| *Regression Statistics* |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Multiple R | 0.9540769  72 |  | | | | | | |
| R Square | 0.9102628  68 |
| Adjusted R  Square | 0.9099989  36 |
| Standard  Error | 630.05959  83 |
| Observatio  ns | 342 |
| ANOVA | | | | | | | | |
|  | *df* | *SS* | *MS* | *F* | *Significan*  *ce F* |  |  |  |
| Regression | 1 | 1.37E+  09 | 1.37E+  09 | 3448.8  44 | 4.6E-180 |  |  |  |
| Residual | 340 | 1.35E+  08 | 396975  .1 |  |  |  |  |  |
| Total | 341 | 1.5E+0  9 |  |  |  |  |  |  |
|  | | | | | | | | |
|  | *Coefficient*  *s* | *Standar*  *d Error* | *t Stat* | *P-value* | *Lower*  *95%* | *Upper*  *95%* | *Lower*  *95.0%* | *Upper*  *95.0%* |
| Intercept | 2068.9931  61 | 88.479  52 | 23.383  87 | 9.14E-  73 | 1894.957 | 2243.0  29 | 1894.9  57 | 2243.0  29 |
| X Variable  1 | 246.46556  83 | 4.1968  12 | 58.726  86 | 4.6E-  180 | 238.2106 | 254.72  06 | 238.21  06 | 254.72  06 |

#### Anova : Single Factor

The single-factor ANOVA conducted on the quantity (Qty) and sales amount (Amount) reveals a significant difference between the groups. The analysis indicates that there's a substantial variance between the groups (SS = $8.01E+09) compared to within the groups (SS =

$1.5E+09), resulting in a high F-statistic of 3632.879 with a very low p-value (nearly zero). This implies that the difference in means between the quantity and sales amount is unlikely to have occurred by chance. Therefore, we reject the null hypothesis and conclude that there's a significant difference in sales amounts attributed to different quantities sold.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Anova: Single Factor | | | | | | |
| SUMMARY | | | | | | |
| *Groups* | *Count* | *Sum* | *Average* | *Variance* |  |  |
| Qty | 342 | 6654.271 | 19.45693 | 66.0952 |  |  |
| Amount | 342 | 2347644 | 6864.457 | 4410782 |  |  |
|  |  |  |  |  |  |  |
| ANOVA | | | | | | |
| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Between Groups | 8.01E+09 | 1 | 8.01E+09 | 3632.879 | 2.1E-275 | 3.85513 |
| Within Groups | 1.5E+09 | 682 | 2205424 |  |  |  |
|  |  |  |  |  |  |  |
| Total | 9.52E+09 | 683 |  |  |  |  |

#### Anova two factor

In the two-factor ANOVA analysis without replication, we observe both rows and columns contribute significantly to the variance. The sums of squares (SS) for rows and columns are

$7.58E+08 and $8.01E+09, respectively. The high F-statistics for both rows (1.014883) and columns (3659.913) with low p-values (nearly zero) indicate that the differences observed in both factors are statistically significant. Therefore, we reject the null hypothesis and conclude that both the quantity sold (Qty) and sales amount (Amount) significantly affect the variance in the dataset.

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Anova: Two-Factor Without Replication | | | | | | | | |  | | |
|  | | | | | | | | |
| *SUMMARY* | *Count* | | *Sum* | | *Average* | | *Variance* | |
| Row 1 | 2 | | 1003 | | 501.5 | | 497004.5 | |
| Row 2 | 2 | | 7804 | | 3902 | | 30388808 | |
| Row 3 | 2 | | 3005 | | 1502.5 | | 4485013 | |
| Row 4 | 2 | | 2304 | | 1152 | | 2635808 | |
| Row 5 | 2 | | 7003 | | 3501.5 | | 24479005 | |
| Row 339 | 2 | | 10252.82 | | 5126.411 | | 51884342 | |
| Row 340 | 2 | | 10272.93 | | 5136.467 | | 52087770 | |
| Row 341 | 2 | | 10293.05 | | 5146.523 | | 52291595 | |
| Row 342 | 2 | | 10313.16 | | 5156.58 | | 52495819 | |
|  | | | | | | | | |
| Qty | 342 | | 6654.271 | | 19.45693 | | 66.0952 | |
| Amount | 342 | | 2347644 | | 6864.457 | | 4410782 | |
| ANOVA | | | | | | | | | | | |
| *Source of Variation* | | *SS* | | *df* | | *MS* | | *F* | | *P-value* | *F crit* |
| Rows | | 7.58E+08 | | 341 | | 2221714 | | 1.014883 | | 0.445792 | 1.195299 |
| Columns | | 8.01E+09 | | 1 | | 8.01E+09 | | 3659.913 | | 2.1E-184 | 3.868873 |
| Error | | 7.46E+08 | | 341 | | 2189134 | |  | |  |  |
|  | |  | |  | |  | |  | |  |  |
| Total | | 9.52E+09 | | 683 | |  | |  | |  |  |

#### Descriptive Statistics:

For the quantity sold (Qty), the mean is approximately 19.46 with a standard error of 0.44. The data is moderately positively skewed (skewness = -0.10) and shows a slight negative kurtosis (-0.999), indicating a slightly flatter distribution compared to a normal distribution. The range of quantity sold spans from 3 to 33.31.

For the sales amount (Amount), the mean is approximately 6864.46 with a larger standard error of 113.57. The data is also moderately positively skewed (skewness = -0.36) and has a slightly negative kurtosis (-0.508). The range of sales amount is much larger, ranging from 1000 to 10279.85.

|  |  |  |  |
| --- | --- | --- | --- |
| *Qty* |  | *Amount* |  |

|  |  |  |  |
| --- | --- | --- | --- |
|  | | | |
| Mean | 19.45693 | Mean | 6864.457 |
| Standard Error | 0.439614 | Standard Error | 113.5651 |
| Median | 19.45693 | Median | 6984.647 |
| Mode | 3 | Mode | 1000 |
| Standard Deviation | 8.129896 | Standard Deviation | 2100.186 |
| Sample Variance | 66.0952 | Sample Variance | 4410782 |
| Kurtosis | -0.99883 | Kurtosis | -0.5078 |
| Skewness | -0.09948 | Skewness | -0.36449 |
| Range | 30.30852 | Range | 9279.851 |
| Minimum | 3 | Minimum | 1000 |
| Maximum | 33.30852 | Maximum | 10279.85 |
| Sum | 6654.271 | Sum | 2347644 |
| Count | 342 | Count | 342 |

#### Correlation

The correlation coefficient between quantity sold (Qty) and sales amount (Amount) is approximately 0.954. This strong positive correlation suggests that there is a significant relationship between the quantity of items sold and the corresponding sales amount, indicating that as the quantity sold increases, the sales amount also tends to increase.

|  |  |  |
| --- | --- | --- |
|  | *Qty* | *Amount* |
| Qty | 1 |  |
| Amount | 0.954077 | 1 |

Introduction

Sales Data Sample Report

This report analyses a comprehensive sales dataset, featuring attributes such as ORDERNUMBER, QUANTITYORDERED, PRICEEACH, and SALES. It aims to extract

insights to guide sales strategies and enhance business performance. The intended audience includes sales managers, marketers, and executives seeking to optimize sales operations and maximize revenue. Key analyses include comparing sales of Vintage cars and Classic cars, determining average sales, identifying top-selling products, assessing profit by country for specific product lines, comparing sales across different years, and evaluating countries based on deal size. Through these analyses, the report aims to provide actionable insights for driving sales growth and improving overall business outcomes.

The scope of the project encompasses analysing a comprehensive sales dataset to extract valuable insights that can inform sales strategies, optimize product offerings, and enhance overall business performance. analysts and researchers seeking insights into sales dynamics and market trends will find value in the project.

### Questionnaire

1. Comparison of sales between Vintage cars and Classic cars across all countries.
2. Determination of the average sales of all products and identification of the highest-selling product.
3. Assessment of the country yielding the most profit for Motorcycles, Trucks, and Buses.
4. Comparison of sales for all items across the years 2004 and 2005.
5. Comparative analysis of all countries based on deal size.

### Analytics

1. Comparison of sales between Vintage cars and Classic cars across all countries.

This analysis Compare the sale of Vintage cars and Classic cars for all the countries. Where USA(2102394.02) has the highest sales followed by Spain, France, and Australia.



**Compare the sale of Vintage cars and Classic cars for all the countries**

2500000

2000000

1500000

1000000

500000

0

382640.86

128656.95

62062.56

102136.01

178288.29

171935.24

565561.01

169250.91

33923.22

239027.39

76721.31

177808.37

55047.18

167850.9

705679.66

102892.52

117713.56

283176.44

2102394.12



1. Determination of the average sales of all products and identification of the highest-selling product.

This analysis aims to provide average sales of all products and identification of the highest- selling product. And through the graph we can see that Classic Cars have the highest sales with 4053.377104 average sales followed by Trucks and Buses and Motorcycles.

Compare the average sales of each product line.

Vintage Cars

3135.33911

Trucks and Buses

3746.8101

Trains

2938.226883

Ships

3053.150128

Planes

3186.286176

Motorcycles

3523.831843

Classic Cars

4053.377104

0

500

1000

1500

2000

2500

3000

3500

4000

4500

1. Assessment of the country yielding the most profit for Motorcycles, Trucks, and Buses.

This analysis aims to identify the country yielding the most profit for Motorcycles, Trucks, and Buses. And bar chart shows that the USA has the highest sales with 397842.42 sum of sales for Trucks and Buses while 520371.7 sum of sales for Motorcycles followed by France and Spain.



Compare the sales of Motorcycles, Trucks, and Buses for each country.

USA

UK

Sweden Spain Singapore Philippines Norway Japan

Italy Ireland Germany France Finland Denmark Canada Austria

Australia

397842.42

520371.7

28142.99

40802.81

15567.25

47931.27

4175.6

74634.82

89027.68

177556.78

18061.68

37075.64

51768.63

13349.44

26536.41

5914.97

Trucks and Buses

7567.8

3983.05

4953.2

Motorcycles

7497.5

10178

116982.22

226390.31

40479.33

47866.72

9588.82

4177.49

51945.98

20472.75

26047.66

77318.5

89968.76

0

100000

200000

300000

400000

500000

600000

1. Comparison of sales for all items across the years 2004 and 2005.

This analysis aims to compare the sales for all the items across the years 2004 and 2005, with the line chart we can see that the sales for all the items across the years are shifting at very rate where the sales for Classic cars are highest among all the categories in both the years with 1762257.09 sales in 2004 and 672573.28 sales in 2005.



2000000

1800000

1600000

1400000

1200000

1000000

Comparison of sales for all items across the years 2004 and

2005

1762257.09

911423.77

672573.28

2004

529302.89

340739.31

2005

3.85

178057.02

.33

Classic Cars Motorcycles Planes

Ships

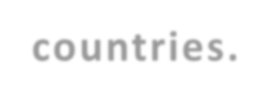
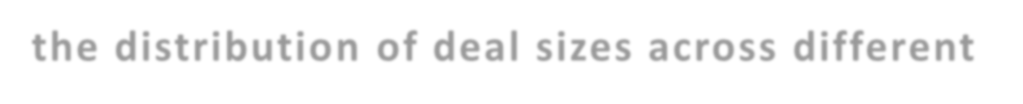
Trains Trucks and Vintage Cars

Buses

|  |  |  |
| --- | --- | --- |
| 800000 |  | 560545.23 502671.8 |
| 600000 |  | 341437.97 |
| 400000 |  | 234947.53 200074.17 128178.07 11652 |
| 200000 |  | 36917 |
| 0 |  |  |

1. Comparative analysis of all countries based on deal size.

This analysis aims to find out the distribution of deal sizes across the different countries. And the bar chart shows that the deal size in the USA with large deal size of 64, medium deal size of 505, and small deal size of 435 is way higher than all the other countries.



**the distribution of deal sizes across different**

**countries.**

USA 64

505

435

UK 4 71

Switzerland 229

Sweden 23124

Spain 17

Singapore 438 37

Philippines 1141

Norway 542 38

Japan 32128

69

171

154

Italy 7 44 62

Ireland 268

Germany 43028

France 21 149 144

Finland 5 46 41

Denmark 63126

Canada 133 36

Belgium 1518

Austria 42922

Australia

7 86

0

92

200

400

600

800

1000

1200

Large Medium Small

### Conclusion and Review

The analysis uncovers significant insights into sales dynamics and profitability across categories and countries. Notably, the USA emerges as a key market leader, exhibiting strong sales performance in Vintage and Classic cars, Trucks, Buses, and Motorcycles. Classic Cars stand out as the highest-selling product, contributing significantly to overall sales revenue. Moreover, the USA demonstrates exceptional profitability, particularly in the Trucks, Buses, and Motorcycles categories. Sales for Classic cars remain consistently robust throughout the years 2004 and 2005, indicating sustained demand for this category. Additionally, the USA showcases markedly larger deal sizes compared to other countries, underscoring its dominance in sales volume.

While the analysis effectively presents key findings through visualizations, further exploration into factors influencing sales fluctuations and deal size disparities could provide deeper insights. Overall, the report offers valuable insights for optimizing sales strategies and driving business growth.

### Regression

This regression analysis for the sales dataset reveals that the model is statistically significant, as indicated by a very low p-value (4.62E-78). The multiple R value of 0.877 suggests a strong positive linear relationship between the independent variables (MSRP, Quantity Ordered) and the dependent variable (Sales). The coefficient values indicate that for every unit increase in MSRP, there's an increase of approximately $103.08 in sales. Similarly, for every unit increase in Quantity Ordered, sales increase by about $12.82, and for every unit increase in the third independent variable, sales increase by approximately $47.43. The adjusted R-squared value of 0.766 indicates that the model explains about 76.6% of the variance in the sales data.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| SUMMARY OUTPUT | |  | | | | | | |
|  | |
| *Regression Statistics* | |
| Multiple R | 0.877178 |
| R Square | 0.769441 |
| Adjusted R  Square | 0.766629 |
| Standard  Error | 896.6688 |
| Observatio  ns | 250 |
| ANOVA | | | | | | | | |
|  | *df* | *SS* | *MS* | *F* | *Significan*  *ce F* |  |  |  |
| Regression | 3 | 6.6E+0  8 | 2.2E+0  8 | 273.65  67 | 4.62E-78 |  |  |  |
| Residual | 246 | 1.98E+  08 | 804014  .9 |  |  |  |  |  |
| Total | 249 | 8.58E+  08 |  |  |  |  |  |  |
|  | | | | | | | | |
|  | *Coefficien*  *ts* | *Standar*  *d Error* | *t Stat* | *P-value* | *Lower*  *95%* | *Upper*  *95%* | *Lower*  *95.0%* | *Upper*  *95.0%* |
| Intercept | -5271.93 | 322.91  66 | -16.326 | 4.32E-  41 | -5907.96 | -4635.9 | - 5907.9  6 | -4635.9 |
| X Variable  1 | 103.0809 | 6.0011  52 | 17.176  85 | 5.42E-  44 | 91.26071 | 114.90  11 | 91.260  71 | 114.90  11 |
| X Variable  2 | 12.81807 | 1.6617  34 | 7.7136  68 | 3.04E-  13 | 9.545024 | 16.091  11 | 9.5450  24 | 16.091  11 |
| X Variable  3 | 47.42944 | 3.3509  38 | 14.154  08 | 1.13E-  33 | 40.82925 | 54.029  63 | 40.829  25 | 54.029  63 |

### Anova: one factor

In this single-factor ANOVA analysis, we're comparing the impact of different levels of the factor (Sales and MSRP) on the variance in the data. The ANOVA results indicate that there is a significant difference between the groups, as the p-value is very low (3.1E-113). This suggests that there's strong evidence to reject the null hypothesis, indicating that at least one of the means of the groups (Sales and MSRP) is significantly different from the others. The F- value of 894.0704 further supports this, as it is much greater than 1, indicating that there is a significant difference between the groups. Therefore, there is evidence to suggest that Sales and MSRP have a significant impact on the variance in the dataset.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Anova: Single Factor | | | | | | |
| SUMMARY | | | | | | |
| *Groups* | *Count* | *Sum* | *Average* | *Variance* |  |  |
| Sales | 250 | 903280.9 | 3613.123 | 3445221 |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| MSRP | 250 | 25534 | 102.136 | 1664.552 |  |  |
| ANOVA | | | | | | |
| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Between Groups | 1.54E+09 | 1 | 1.54E+09 | 894.0704 | 3.1E-113 | 3.860199 |
| Within Groups | 8.58E+08 | 498 | 1723443 |  |  |  |
|  |  |  |  |  |  |  |
| Total | 2.4E+09 | 499 |  |  |  |  |

### Anova: two factor

This two-factor ANOVA without replication analyzes the impact of Sales, MSRP, and Quantity Ordered on the variance in the dataset. The ANOVA results show that there is no significant difference between the rows (Sales, MSRP, and Quantity Ordered) as the p-value (0.33951) is greater than the significance level (0.05). However, there is a significant difference between the columns (Sales and MSRP) with a very low p-value (1.9E-168), indicating that at least one of the means of the groups is significantly different from the others. The F-value of 925.2361 further supports this conclusion. Therefore, we can infer that Sales and MSRP have a significant impact on the variance in the dataset.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Anova: Two-Factor Without Replication | | | | | |  | |
|  | | | | | |
| *SUMMARY* | *Count* | | *Sum* | *Average* | *Variance* |
| Row 1 | 3 | | 4097.66 | 1365.887 | 5069957 |
| Row 2 | 3 | | 2451.12 | 817.04 | 1725170 |
| Row 3 | | 3 | 1566 | 522 | 648687 |
| Row 4 | | 3 | 5095.24 | 1698.413 | 7507173 |
| Row 5 | | 3 | 5140.39 | 1713.463 | 7650609 |
| Row 248 | | 3 | 4386.35 | 1462.117 | 5944534 |
| Row 249 | | 3 | 2261.6 | 753.8667 | 1546167 |
| Row 250 | | 3 | 4176.72 | 1392.24 | 5420980 |
| Sales | | 250 | 903280.9 | 3613.123 | 3445221 |
| MSRP | | 250 | 25534 | 102.136 | 1664.552 |
| QuantityOrdered | | 250 | 8659 | 34.636 | 89.69428 |
| ANOVA | | | | | | | |
| *Source of Variation* | | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Rows | | 2.95E+08 | 249 | 1182944 | 1.044989 | 0.33951 | 1.194432 |
| Columns | | 2.09E+09 | 2 | 1.05E+09 | 925.2361 | 1.9E-168 | 3.013826 |
| Error | | 5.64E+08 | 498 | 1132016 |  |  |  |
|  | |  |  |  |  |  |  |
| Total | | 2.95E+09 | 749 |  |  |  |  |
|  | |  |  |  |  |  |  |

### Descriptive Statistics

The descriptive statistics for Quantity Ordered, Sales, MSRP, and Price Each reveal valuable insights into the dataset. Quantity Ordered has a mean of 34.636 units, with a standard deviation of 9.470706, indicating moderate variability in the quantity ordered. Sales, on the other hand, show a much higher variability, with a mean of 3613.123 and a standard deviation of 1856.131. The MSRP (Manufacturer's Suggested Retail Price) has a mean of 102.136, with a standard

deviation of 40.79892, suggesting moderate variability in the price. In contrast, Price Each, with a mean of 84.45296 and a standard deviation of 20.22993, exhibits less variability compared to MSRP. The skewness and kurtosis values provide insights into the distribution shape and tail behaviour of the variables. Overall, these descriptive statistics offer a comprehensive understanding of the dataset's central tendency, variability, and distribution characteristics for each variable.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| *Quantity*  *Ordered* |  | *Sales* |  | *MSRP* |  | *Price Each* |  |
|  | | | | | | | |
| Mean | 34.636 | Mean | 3613.12  3 | Mean | 102.136 | Mean | 84.4529  6 |
| Standard  Error | 0.59898 | Standard  Error | 117.392 | Standard  Error | 2.58035 | Standard  Error | 1.27945  3 |
| Median | 34 | Median | 3263.96 | Median | 99 | Median | 100 |
| Mode | 29 | Mode | #N/A | Mode | 118 | Mode | 100 |
| Standard  Deviation | 9.47070  6 | Standard  Deviation | 1856.13  1 | Standard  Deviation | 40.7989  2 | Standard  Deviation | 20.2299  3 |
| Sample  Variance | 89.6942  8 | Sample  Variance | 3445221 | Sample  Variance | 1664.55  2 | Sample  Variance | 409.249  9 |
| Kurtosis | -0.64676 | Kurtosis | 1.12705  7 | Kurtosis | -0.19836 | Kurtosis | -0.40344 |
| Skewness | 0.25674  5 | Skewness | 1.01348  9 | Skewness | 0.51710  4 | Skewness | -0.9678 |
| Range | 51 | Range | 10626.8  5 | Range | 181 | Range | 73.12 |
| Minimum | 15 | Minimum | 652.35 | Minimum | 33 | Minimum | 26.88 |
| Maximum | 66 | Maximum | 11279.2 | Maximum | 214 | Maximum | 100 |
| Sum | 8659 | Sum | 903280.  9 | Sum | 25534 | Sum | 21113.2  4 |
| Count | 250 | Count | 250 | Count | 250 | Count | 250 |

### Correlation

The correlation matrix indicates the relationships between Quantity Ordered, Sales, and Price Each. There's a moderate positive correlation of approximately 0.514 between Quantity Ordered and Sales, suggesting that as the quantity ordered increases, sales tend to increase as well. Similarly, there's a weak positive correlation of about 0.664 between Sales and Price Each, indicating that higher-priced items may contribute to higher sales, albeit to a lesser extent. However, there seems to be a negligible correlation of approximately -0.013 between Quantity Ordered and Price Each, implying that changes in the quantity ordered don't significantly impact the individual item price.

|  |  |  |  |
| --- | --- | --- | --- |
|  | *Quantity Ordered* | *Sales* | *Price Each* |
| Quantity Ordered | 1 |  |  |
| Sales | 0.513951 | 1 |  |
| Price Each | -0.01254 | 0.663973 | 1 |

Introduction

## Store Dataset Report

This dataset comprises sales data from a retail store, encompassing various attributes such as customer demographics (Gender, Age Group), transaction details (Order ID, Status), product specifics (Category, SKU), and shipping information. Our analysis is geared towards elucidating customer behavior and product trends, with the goal of uncovering patterns, preferences, and correlations within the data. By harnessing these insights, businesses can refine marketing strategies, streamline inventory management, and elevate overall customer satisfaction.

### Questionnaire

1. Compare various channels based on how many male customers order and female customer order.
2. Compare all the categories of order where amount is less than 1500 and greater than 5000.
3. How many Customers are there whose age is 30 and above and state is Delhi.
4. Which of the following state perform better than other, Delhi, Tamil Nadu, Maharashtra, Rajasthan.
5. Which city performed better than all other cities based on highest order placed.
6. Compare various categories of items based on most quantity sold and show which gender buys the most category.

### Analytics

1. Compare various channels based on how many male customers order and female customer order?

Amazon leads in the sales in both men and women category followed by Myntra and Flipkart. Amazon sold almost 3432 units in men category and almost 7547 units in women category. Myntra sold 2156 units in men section and 5062 units in women section.

Compare various channels based on the number of male and

female customers' orders

Others

393 864

Nalli

454

1026

Myntra

2156

5062

Meesho

395

993

Women

Flipkart

2043

4643

Men

Amazon

3432

7547

Ajio

579

1344

0

1000

2000

3000

4000

5000

6000

7000

8000

1. Compare all the categories of order where amount is less than 1500 and greater than 5000. This analysis helps in comparing the categories of order where amount is less than 1500 and

greater than 5000. Showing the kurta(12391) and set(10446) with highest count of the orders followed by western dress, top and saree.

Compare all categories of orders where the amount is less

than 1500 and greater than 5000

14000

12391

12000

10446

10000

8000

6000

4066

4000

2193

2000

1380

229 78 264

0

Blouse

Bottom Ethnic Dress kurta

Saree

Set

Top

Western

Dress

1. Which of the following state perform better than other, Delhi, Tamil Nadu, Maharashtra, Rajasthan.

This analysis shows which states performed better than the states mentioned above and Karnataka (2646358) has the highest performance than the other states followed by Uttar Pradesh(2104659).

Compare the performance of Delhi, Tamil Nadu,

Maharashtra, and Rajasthan

UTTARAKHAND

TRIPURA TAMIL NADU RAJASTHAN PUDUCHERRY

New Delhi MIZORAM MANIPUR MADHYA PRADESH

KERALA JHARKHAND HIMACHAL PRADESH

GUJARAT

DELHI CHHATTISGARH

BIHAR ARUNACHAL PRADESH

ANDAMAN & NICOBAR

327179

922444

30961

2104659

54916

1712439

1678877

547360

368940

48553

414840

8422

43510

12182

25988

78865

14148

564026

2990221

1008940

255054

158736

146246

2646358

184169

14980

174531

813320

715563

1266328

63059

36840

51970

446831

326423

918499

0

500000 1000000 1500000 2000000 2500000 3000000 3500000

1. Which city performed better than all other cities based on highest order placed.

Based on the graph recorded we can actually see which city performed better than all other cities based on highest order placed, so according to graph Bangluru has the highest order placed with 2673 orders followed by Hyderabad(1998).



the city that performed better than all others based on the highest order placed

1600

1400

1200

1000

800

600

400

200

0

1. Compare various categories of items based on most quantity sold and also show which gender buys the most category.

This analysis shows the comparison of various categories of items based on most quantity sold which is kurta bought by women set bought by women followed by men and western dress followed by top for both men and women.

Compare various categories of items based on the most

quantity sold and show which gender buys the most category

Western Dress

459

3606

Top

49

2144

Set

7260

Saree

289

1088

kurta

174

10306

Women

Men

Ethnic Dress

88

176

Bottom

32

46

Blouse

225

7

0

2000

4000

6000

8000

10000

12000

5139

### Conclusion And Review

The analysis highlights Amazon's dominance in sales across both men's and women's categories, with Myntra and Flipkart following closely behind. Amazon leads in sales for both men's and women's categories, followed by Myntra and Flipkart. Top-selling items include kurta and set, with Karnataka and Bangalore showing the highest sales performance.

The analysis provides valuable insights into sales trends and regional performance, aiding decision-making for retailers. However, further exploration into additional factors influencing sales could enhance the analysis. Overall, the findings offer valuable information for optimizing sales strategies in competitive markets.

### Regression

The regression analysis for the store dataset indicates that the model has a multiple ( R ) value of approximately 0.172, suggesting a weak positive correlation between the independent variables (quantity and size) and the dependent variable (amount). The ( R^2 ) value, which measures the proportion of the variance in the dependent variable explained by the independent variables, is approximately 0.030. This suggests that only about 3% of the variability in the amount can be explained by the quantity and size.

In terms of significance, the ANOVA results show that the regression model is statistically significant, as indicated by the very low p-value of 0 for the regression. However, when looking at the coefficients, it's observed that the coefficient for X Variable 1 (quantity) is not statistically significant, with a p-value of 0.632. On the other hand, the coefficient for X Variable 2 (size) is highly significant, with a very low p-value (approximately 1.3 \* 10^{-205}), indicating that size has a substantial impact on the amount.

The intercept term is also statistically significant, indicating that even when quantity and size are zero, there is still a significant amount expected, as evidenced by the low p-value and the confidence intervals for the intercept coefficient.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| SUMMARY OUTPUT | |  | | | |
| *Regression Statistics* | |
| Multiple R | 0.172398 |
| R Square | 0.029721 |
| Adjusted R  Square | 0.029659 |
| Standard Error | 264.5693 |
| Observations | 31047 |
|  | |
| ANOVA | | | | | |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |
| Regression | 2 | 66561870 | 33280935 | 475.4629 | 0 |
| Residual | 31044 | 2.17E+09 | 69996.92 |  |  |
| Total | 31046 | 2.24E+09 |  |  |  |
|  | | | | | |
|  | *Coefficients* | *Standard*  *Error* | *t Stat* | *P-value* | *Lower 95%* |
| Intercept | 185.155 | 16.57854 | 11.16836 | 6.61E-29 | 152.6604 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| X Variable 1 | 0.047626 | 0.099327 | 0.479489 | 0.631594 | -0.14706 |
| X Variable 2 | 492.0276 | 15.95904 | 30.83065 | 1.3E-205 | 460.7472 |

### Anova-1 factor

The single-factor ANOVA test conducted on the Qty and Amount groups reveals a highly significant result. The between-groups variance, which measures the variability between the Qty and Amount groups, is extremely large (SS = 7.2 \* 10^9), resulting in a very high F-statistic (F = 199639.8) and an associated p-value close to zero (p < 0.001). This indicates that there is a significant difference between the Qty and Amount groups concerning their means. The within-groups varian.ce, representing the variability within each group, is also considerable (SS

=2.24 \* 10^9), reflecting the dispersion of data points around their respective group means. Overall, the ANOVA test suggests a strong statistical significance in the difference between Qty and Amount groups.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Anova: Single Factor | | | | | | |
| SUMMARY | | | | | | |
| *Groups* | *Count* | *Sum* | *Average* | *Variance* |  |  |
| Qty | 31047 | 31237 | 1.00612 | 0.008853 |  |  |
| Amount | 31047 | 21176377 | 682.0748 | 72136.38 |  |  |
| ANOVA | | | | | | |
| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Between Groups | 7.2E+09 | 1 | 7.2E+09 | 199639.8 | 0 | 3.841609 |
| Within Groups | 2.24E+09 | 62092 | 36068.2 |  |  |  |
|  | | | | | | |
| Total | 9.44E+09 | 62093 |  |  |  |  |

### Anova- 2 factor

The two-factor ANOVA conducted on Age, Qty, and Amount reveals interesting insights. Regarding rows, the variability in the data across different groups (SS = 7.49 \* 10^8) doesn't show a significant difference between them (p = 0.468). However, the variability between columns is substantial (SS =9.09\*10^9), indicating a significant difference among the factors Age, Qty, and Amount (p < 0.001). The error term, representing variability within groups, is also noteworthy (SS = 1.5\*10^9), showing dispersion within each combination of factors. Overall, the ANOVA results suggest a statistically significant difference between the factors Qty and Amount concerning their means, but no significant difference across age groups.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Anova: Two-Factor Without Replication | | | | |
|  | | | | |
| *SUMMARY* | *Count* | *Sum* | *Average* | *Variance* |
| Row 1 | 3 | 421 | 140.3333 | 42116.33 |
| Row 2 | 3 | 1479 | 493 | 685648 |
| Row 3 | 3 | 521 | 173.6667 | 59609.33 |
| Row 4 | 3 | 750 | 250 | 172171 |
| Row 5 | 3 | 607 | 202.3333 | 88482.33 |
| Row 31044 | 3 | 974 | 324.6667 | 283326.3 |

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Row 31045 | 3 | | 1145 | | 381.6667 | | 403529.3 | |  | | |
| Row 31046 | 3 | | 446 | | 148.6667 | | 47506.33 | |
| Row 31047 | 3 | | 828 | | 276 | | 199225 | |
|  | | | | | | | | |
| Age | 31047 | | 1226250 | | 39.49657 | | 228.5307 | |
| Qty | 31047 | | 31237 | | 1.00612 | | 0.008853 | |
| Amount | 31047 | | 21176377 | | 682.0748 | | 72136.38 | |
| ANOVA | | | | | | | | | | | |
| *Source of Variation* | | *SS* | | *df* | | *MS* | | *F* | | *P-value* | *F crit* |
| Rows | | 7.49E+08 | | 31046 | | 24134.08 | | 1.000774 | | 0.468198 | 1.016275 |
| Columns | | 9.09E+09 | | 2 | | 4.54E+09 | | 188446.6 | | 0 | 2.995877 |
| Error | | 1.5E+09 | | 62092 | | 24115.42 | |  | |  |  |
|  | | | | | | | | | | | |
| Total | | 1.13E+10 | | 93140 | |  | |  | |  |  |

### Descriptive Statistics

The mean age is approximately 39.50 years, with a standard deviation of 15.12. The data's distribution is slightly skewed to the right (skewness = 0.73), and its kurtosis indicates a relatively normal distribution (kurtosis = -0.16). On average, the quantity ordered is about 1.01, with a small standard deviation of 0.09. The mode is 1, indicating that this value appears most frequently in the dataset. However, the data is heavily right-skewed (skewness = 19.45) and exhibits high kurtosis (kurtosis = 475.36), suggesting a heavily tailed distribution. The average amount is approximately 682.07, with a considerable standard deviation of 268.58. The data is moderately skewed to the right (skewness = 1.05) and has a slightly heavier tail (kurtosis = 1.77). The range of values spans from 229 to 3036.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| *Age* |  | *Qty* |  | *Amount* |  |
|  | | | | | |
| Mean | 39.49657 | Mean | 1.00612 | Mean | 682.0748 |
| Standard Error | 0.085795 | Standard Error | 0.000534 | Standard Error | 1.524289 |
| Median | 37 | Median | 1 | Median | 646 |
| Mode | 28 | Mode | 1 | Mode | 399 |
| Standard  Deviation | 15.11723 | Standard  Deviation | 0.094088 | Standard  Deviation | 268.5822 |
| Sample Variance | 228.5307 | Sample Variance | 0.008853 | Sample Variance | 72136.38 |
| Kurtosis | -0.1587 | Kurtosis | 475.3566 | Kurtosis | 1.768676 |
| Skewness | 0.72916 | Skewness | 19.4509 | Skewness | 1.052904 |
| Range | 60 | Range | 4 | Range | 2807 |
| Minimum | 18 | Minimum | 1 | Minimum | 229 |
| Maximum | 78 | Maximum | 5 | Maximum | 3036 |
| Sum | 1226250 | Sum | 31237 | Sum | 21176377 |
| Count | 31047 | Count | 31047 | Count | 31047 |

### Correlation

The correlation matrix reveals the relationships between Age, Qty (quantity), and Amount variables. Firstly, Age shows an almost negligible positive correlation with Qty, with a correlation coefficient of around 0.0049, indicating an extremely weak association. Similarly, the correlation between Age and Amount is also very weak, standing at approximately 0.0035. Conversely, there appears to be a slightly stronger positive correlation, though still weak, between Qty and Amount, with a correlation coefficient of about 0.1724. This suggests that as the quantity ordered increases, there's a modest increase in the total amount. Overall, these correlation values signify subtle connections between the variables, with the quantity ordered having the most notable influence on the total amount compared to Age.

|  |  |  |  |
| --- | --- | --- | --- |
|  | *Age* | *Qty* | *Amount* |
| Age | 1 |  |  |
| Qty | 0.004884 | 1 |  |
| Amount | 0.003522 | 0.172377 | 1 |

## Forecasting

The Shares dataset of Airport showing the time series forecasting of shares from year 2016 to 2050 having the known shares of date series from 2009 to 2015.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Date** | **Airport Shares** | **Forecast(Airpor t Shares)** | **Lower Confidence Bound(Airport**  **Shares)** | **Upper Confidence Bound(Airport**  **Shares)** |
| Jan-09 | 26,44,539 |  |  |  |
| Feb-09 | 23,59,800 |  |  |  |
| Mar-09 | 29,25,918 |  |  |  |
| Apr-09 | 30,24,973 |  |  |  |
| Nov-09 | 29,71,484 |  |  |  |
| Jan-10 | 27,85,466 |  |  |  |
| Feb-10 | 25,15,361 |  |  |  |
| Mar-10 | 31,05,958 |  |  |  |
| Nov-10 | 31,63,659 |  |  |  |
| Jan-11 | 28,83,810 |  |  |  |
| Feb-11 | 26,10,667 |  |  |  |
| Mar-11 | 31,29,205 |  |  |  |
| Apr-11 | 32,00,527 |  |  |  |
| Apr-12 | 35,63,007 |  |  |  |
| May-12 | 38,20,570 |  |  |  |
| Oct-12 | 38,44,987 |  |  |  |
| Nov-12 | 34,78,890 |  |  |  |
| Dec-12 | 34,43,039 |  |  |  |
| Jan-13 | 32,04,637 |  |  |  |
| Feb-13 | 29,66,477 |  |  |  |
| Jul-13 | 41,76,486 |  |  |  |
| Aug-13 | 43,47,059 |  |  |  |
| Sep-13 | 37,81,168 | 37,81,168 | 37,81,168 | 37,81,168 |
| Oct-13 |  |  | -1,62,369 |  |
| Nov-13 |  | 35,62,680 | 33,95,234 | 37,30,125 |
| Dec-13 |  | 36,33,798 | 34,61,388 | 38,06,209 |
| Jan-14 |  | 33,66,457 | 31,89,182 | 35,43,732 |
| Feb-14 |  | 31,10,903 | 29,28,857 | 32,92,949 |
| Apr-15 |  | 38,39,280 | 35,97,219 | 40,81,340 |
| Sep-15 |  | 41,26,863 | 38,65,432 | 43,88,294 |
| Sep-20 |  | 49,93,668 | 45,29,269 | 54,58,066 |
| Sep-25 |  | 58,60,472 | 52,07,346 | 65,13,598 |
| Sep-30 |  | 67,27,277 | 58,83,194 | 75,71,361 |
| Sep-35 |  | 75,94,082 | 65,52,552 | 86,35,611 |
| Sep-40 |  | 84,60,887 | 72,13,929 | 97,07,844 |
| Sep-45 |  | 93,27,691 | 78,66,787 | 1,07,88,596 |
| Sep-50 |  | 1,01,94,496 | 85,10,985 | 1,18,78,007 |

The forecast sheet provides predicted values for Airport Shares from January 2009 to September 2015, with additional projections for September 2020, 2025, 2030, 2035, 2040, 2045 and 2050. These

forecasts are complemented by Lower and Upper Confidence Bounds, indicating the range of expected values with a certain level of confidence. The data allows for insight into expected trends and potential variability in Airport Shares over the specified time period.

The forecasted Airport Shares span a wide range of values across various years, providing insight into the expected trends in airport traffic. For instance, the predicted shares for September 2020, 2025, 2030, 2035, 2040, and 2045 are 49,93,668; 58,60,472; 67,27,277; 75,94,082; 84,60,887; and

93,27,691, respectively and for the 2050 it’s 1,01,94,496. These values indicate a steady increase in airport shares over the years, suggesting a sustained growth trajectory in air travel demand. Additionally, the Lower and Upper Confidence Bounds offer a range of potential values around these predictions, highlighting the uncertainty inherent in forecasting future trends in airport shares.

Forecasting of Airport Shares

14,000,000

12,000,000

10,000,000

8,000,000

6,000,000

4,000,000

2,000,000

0

-2,000,000

Airport Shares

Forecast(Airport Shares)

Lower Confidence Bound(Airport Shares) Upper Confidence Bound(Airport Shares)

The Graph shows the Airport shares, forecast, lower bound, upper bound. The line graph visually represents the variation in Airport Shares over time. The graph illustrates a clear upward trend in airport shares from 2009 to 2050, with occasional fluctuations reflecting changes in air travel demand. Particularly notable is the significant increase in shares projected for the coming decades, indicating a robust growth trajectory in airport traffic. This visual representation effectively communicates the expected trends in airport shares, showcasing the anticipated expansion in air travel activities over the forecasted period.

Jan-09

Apr-09 Jul-09 Oct-09 Jan-10 Apr-10 Jul-10 Oct-10 Jan-11 Apr-11 Jul-11 Oct-11 Jan-12 Apr-12 Jul-12 Oct-12 Jan-13 Apr-13 Jul-13 Oct-13 Jan-14 Apr-14 Jul-14 Oct-14 Jan-15 Apr-15 Jul-15 Sep-20 Sep-35

Sep-50

A compelling narrative of the evolution of airport shares over the specified timeline, offering insights into the dynamic nature of the aviation industry. As depicted, the consistent upward trajectory in airport shares underscores the sector's resilience and enduring appeal, despite occasional fluctuations attributed to various factors such as economic conditions, technological advancements, and geopolitical events. This graphical representation serves as a valuable tool for stakeholders and decision-makers, enabling them to glean actionable insights into the future trajectory of air travel and make informed strategic decisions to capitalize on emerging opportunities and potential challenges.