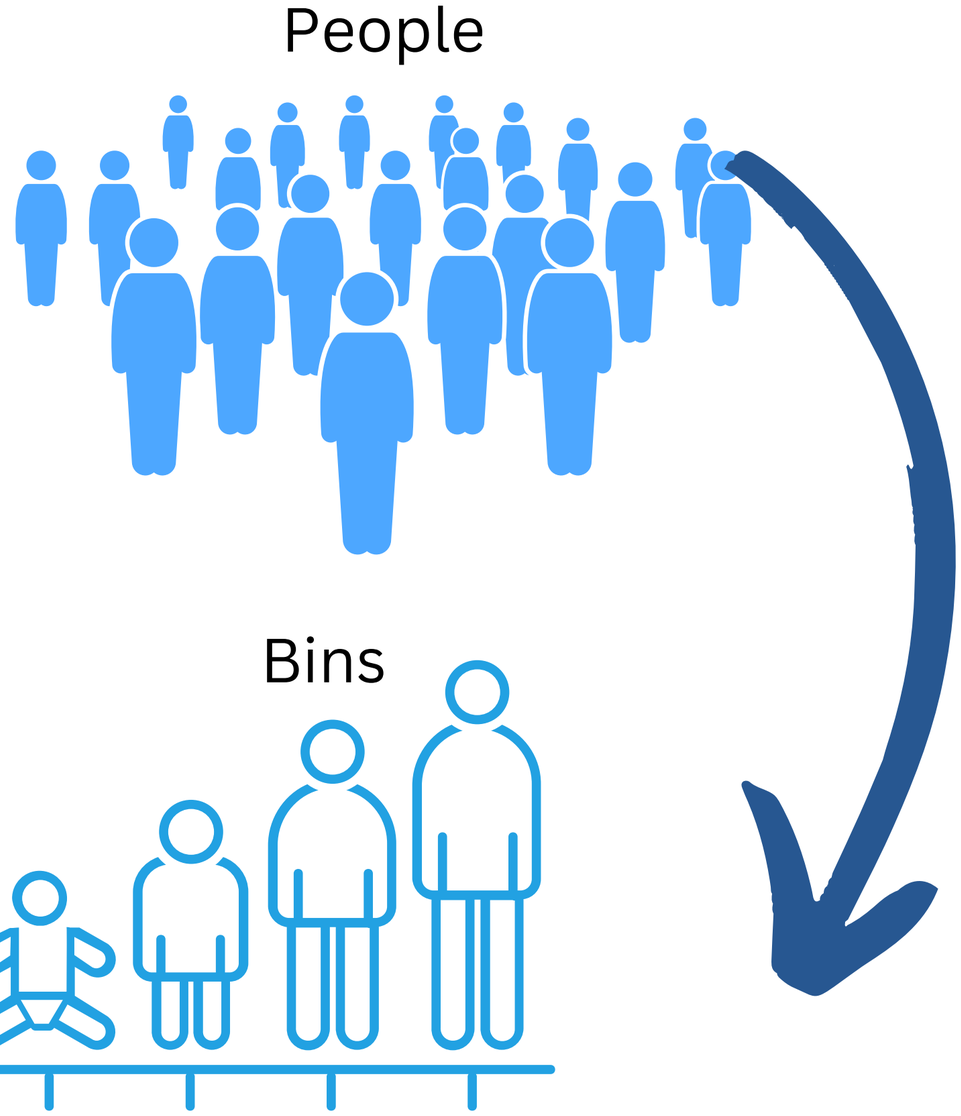
Data Binning

- It's a technique used to convert from a *Numeric* data type into a *Dimension* data type.
- This can be implemented by creating fixed-width bins that group data values inside them.
- Very useful in the case of histograms.

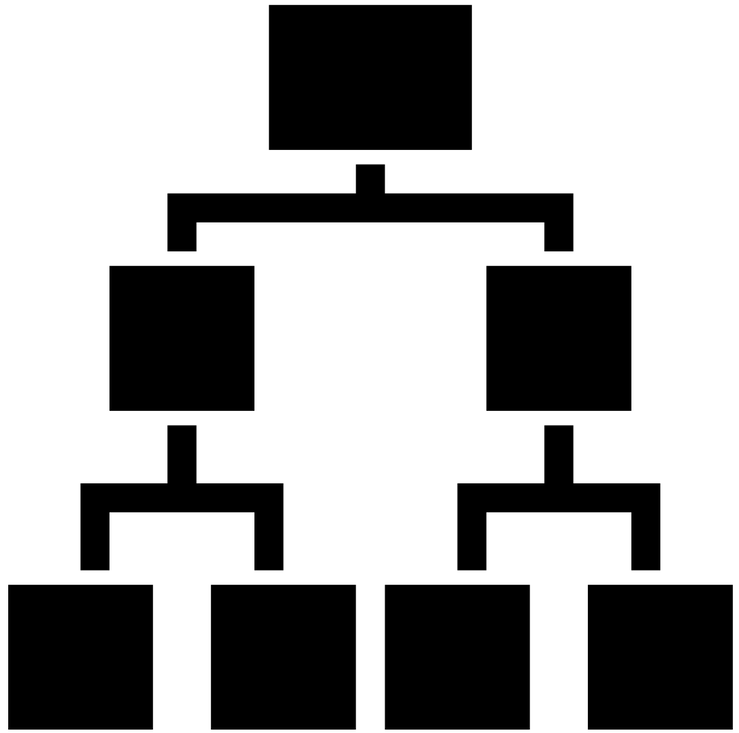
Data Binning

- The age variable is of a numeric data type.
- But in some cases, we might want to have Age Groups instead (Dimension data type).
- We then apply bins with histogram
- of 5 years for example.

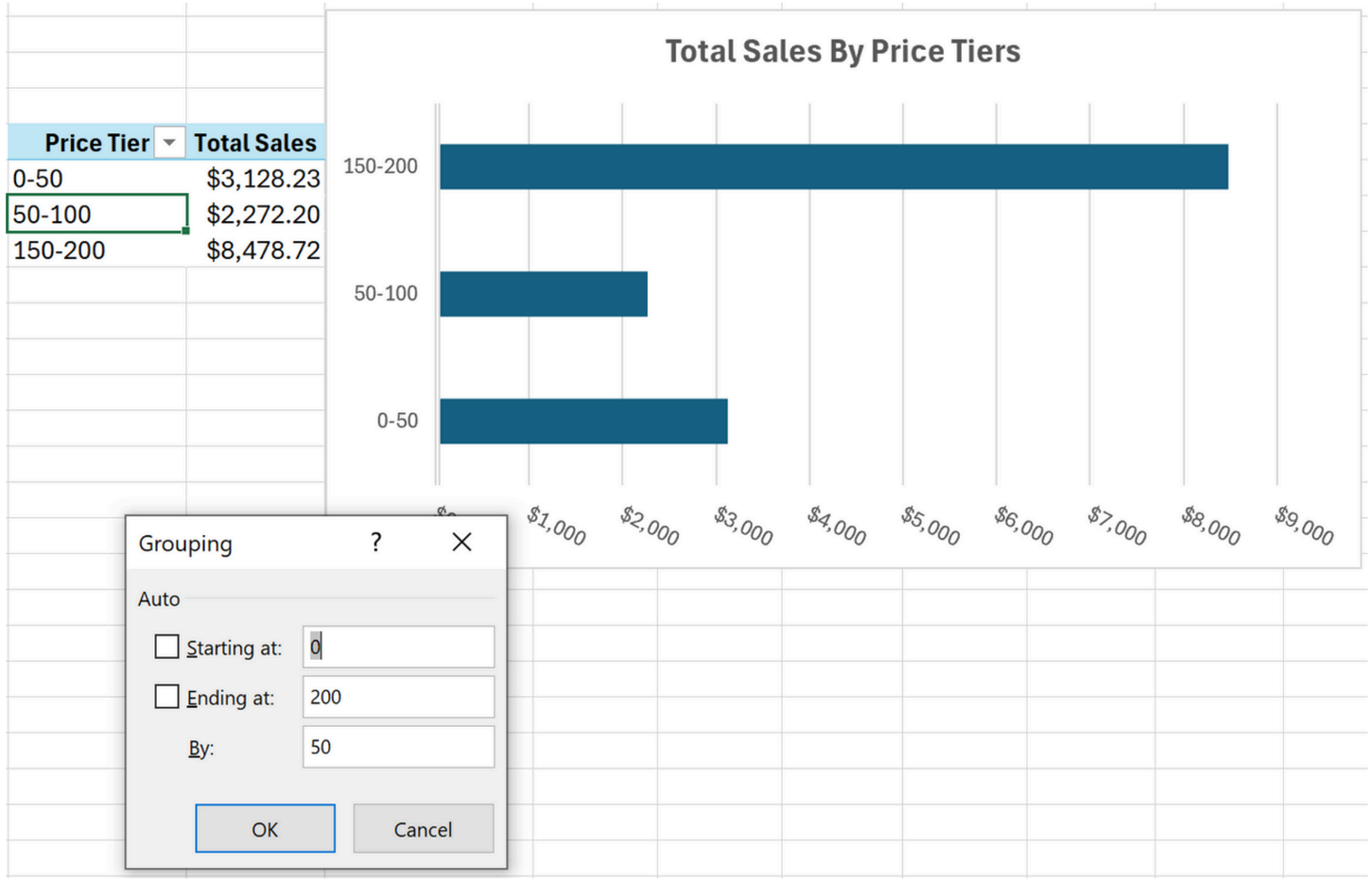


Data Grouping

- It's a technique used to create hierarchical dimensions based on one dimension variable.
-
- The most common examples are the grouping of days into months and months into quarters, ...etc.
-
- It's very useful in the case of data consolidation among these hierarchical dimensions.



Data Binning: Converting Price into Price Tiers



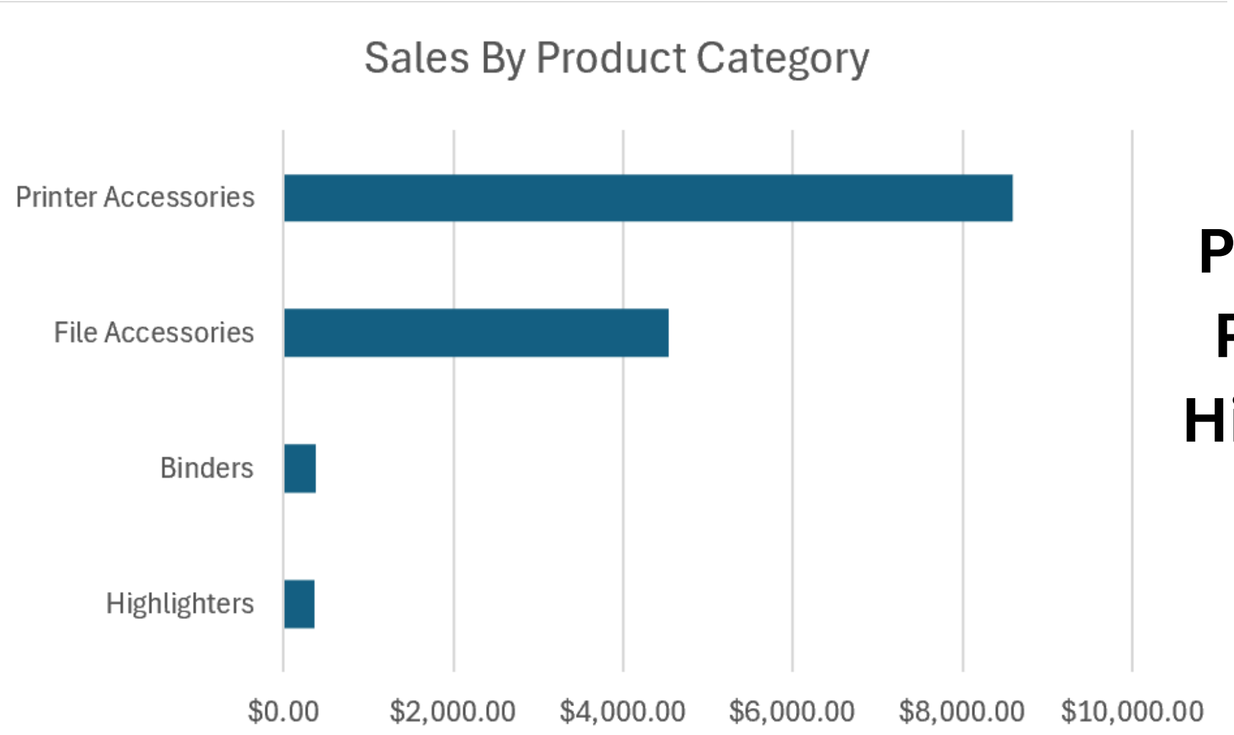
Price --> Measure

Price Tier --> Dimension

Data Grouping

Product Category	Total Sales
⊞ Highlighters	$360.92
⊞ Binders	$387.84
⊞ File Accessories	$4,544.90
⊞ Printer Accessories	$8,585.49

Note the Plus Sign (drill down)



Product --> Dimension

Product Category --> Hierarchical Dimension

Drill Down

Drill down refers to the process of progressively uncovering more specific details within a larger dataset by navigating a hierarchy, where you start with a broad overview and then delve deeper into progressively finer levels of information.

Yearly Sales
Quarterly Sales
Monthly Sales
Daily Sales

Country Level
Regions
Districts
Cities

Roll Up

Roll up is the opposite of drilling down. It refers to the process of summarizing or aggregating data from a more detailed level to a higher level of abstraction. It's like climbing up a hierarchy, where you start with individual data points and progressively combine them into broader categories.

Yearly Sales
Quarterly Sales
Monthly Sales
Daily Sales

Country Level
Regions
Districts
Cities

Table Calculations

In Excel pivot tables, table calculations refer to calculations performed within the pivot table itself, rather than on the underlying source data. This allows you to analyze your data from different perspectives without modifying the original data set.

- Calculating percentages, ratios, and variances
- Creating running totals or averages
- Ranking or filtering data based on specific criteria

Table Calculations

Year	Total Sales	YoY Growth Sales
2013	$11,195.49	
2014	$2,683.66	-76.03%

Show Values As: % Difference From...

YOY Growth = (Sales(this year) - Sales (last year)) / Sales (last year)

Measures

Table Calculations

Year	Total Sales	Running Total Sales
2013	$11,195.49	$11,195.49
2014	$2,683.66	$13,879.15

Show Values As: Running Total In...

Measures

Excel - Charts

Next Page

Categories of Charts

- + RAW NUMBERS
- + TABLES
- + CHANGE OVER TIME
- + DISTRIBUTIONS
- + COMPARISONS
- + RELATIONSHIPS
- + COMPSITIONS

Next Page

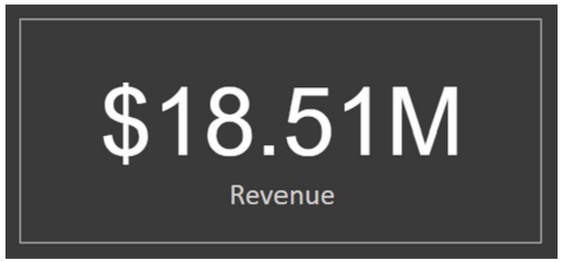
Categories of Charts

+ RAW NUMBERS

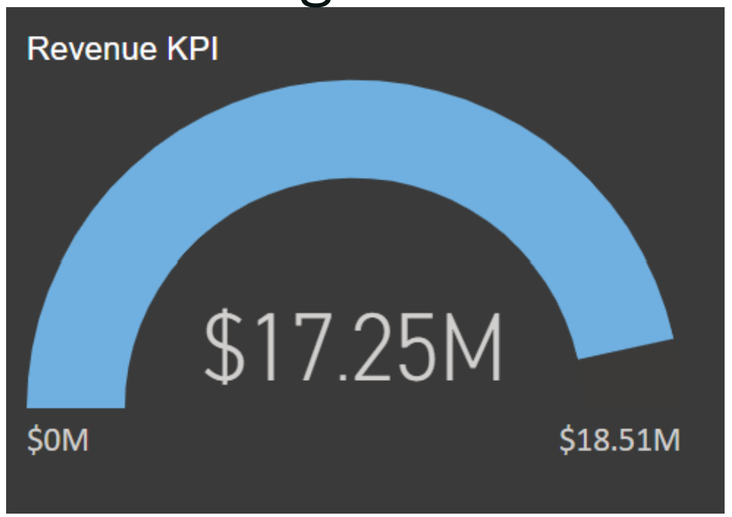
Measures: 1
Dimensions: 0

Bullet Chart

Card



Gauge Chart



Next Page

Tables

Categories of Charts

+ TABLES

Measures: 1 or more
Dimensions: 1

SubcategoryName	Total Profit
Bike Racks	$22,686.24
Bike Stands	$23,290.956
Bottles and Cages	$67,447.7282
Caps	$12,199.789
Cleaners	$8,490.2502
Fenders	$54,487.62
Gloves	$36,578.418
Helmets	$131,036.4003
Hydration Packs	$23,924.4715
Jerseys	$44,956.2461
Mountain Bikes	$3,242,075.7281
Road Bikes	$2,455,330.1642
Shorts	$41,360.1328
Socks	$5,982.2451
Tires and Tubes	$238,396.3948
Touring Bikes	$1,427,160.0465
Vests	$20,710.271
Total	**$7,856,113.1018**

SubcategoryName	Total Cost	Total Profit
Bike Racks	$13,553.76	$22,686.24
Bike Stands	$13,915.044	$23,290.956
Bottles and Cages	$40,297.2118	$67,447.7282
Caps	$23,682.2852	$12,199.789
Cleaners	$5,072.4498	$8,490.2502
Fenders	$32,553.18	$54,487.62
Gloves	$25,682.7584	$36,578.418
Helmets	$74,797.5367	$131,036.4003
Hydration Packs	$14,293.5785	$23,924.4715
Jerseys	$113,608.6904	$44,956.2461
Mountain Bikes	$3,809,222.2937	$3,242,075.7281
Road Bikes	$3,959,368.6689	$2,455,330.1642
Shorts	$24,710.4272	$41,360.1328
Socks	$3,574.1249	$5,982.2451
Tires and Tubes	$142,430.2852	$238,396.3948
Touring Bikes	$2,344,404.6135	$1,427,160.0465
Vests	$12,373.229	$20,710.271
Total	**$10,653,540.1372**	**$7,856,113.1018**

Next Page

Tables

Categories of Charts

+ TABLES

Measures: 1 or more
Dimensions: 1 or more

Matrix

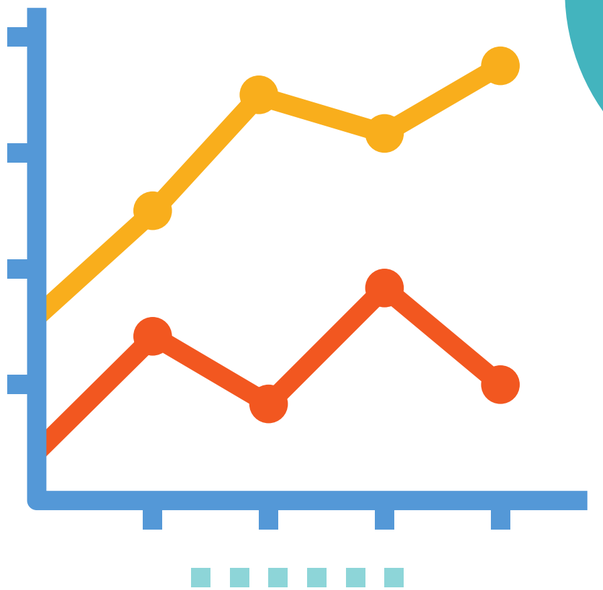
Year	Accessories	Bikes	Clothing
2016	$250,936.123	$3,647,021.0155	$69,126.9885
2017	$318,823.938	$3,477,544.9233	$92,660.1135

Next Page

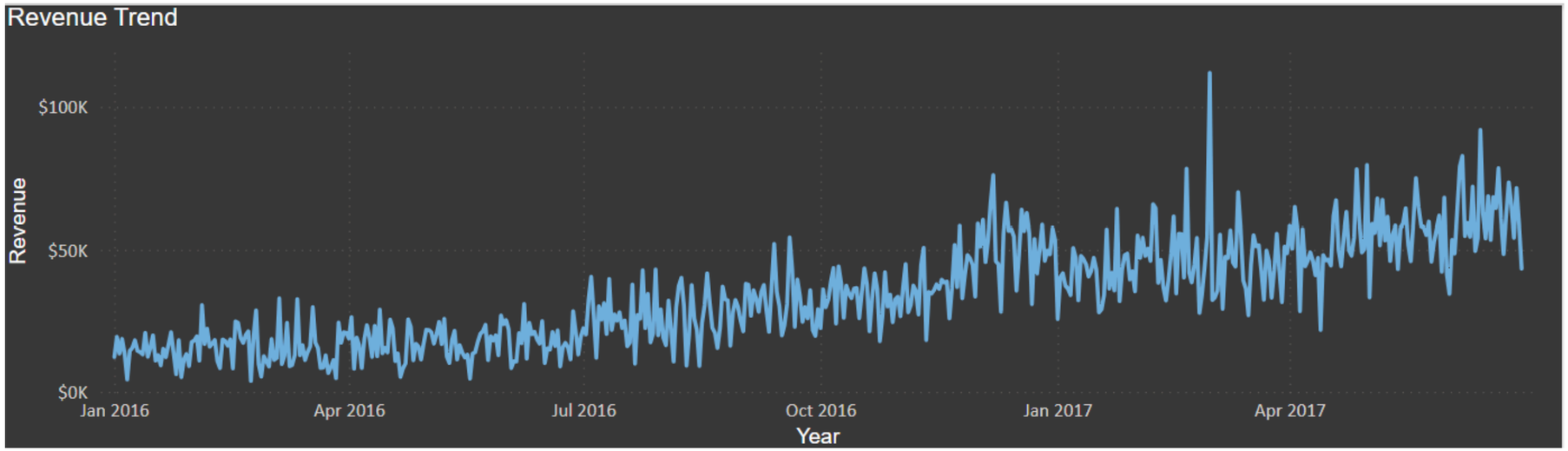
Categories of Charts

+ CHANGE OVER TIME

Measures: 1 or more
Dimensions: 1 (Continuous Time)



Line Chart



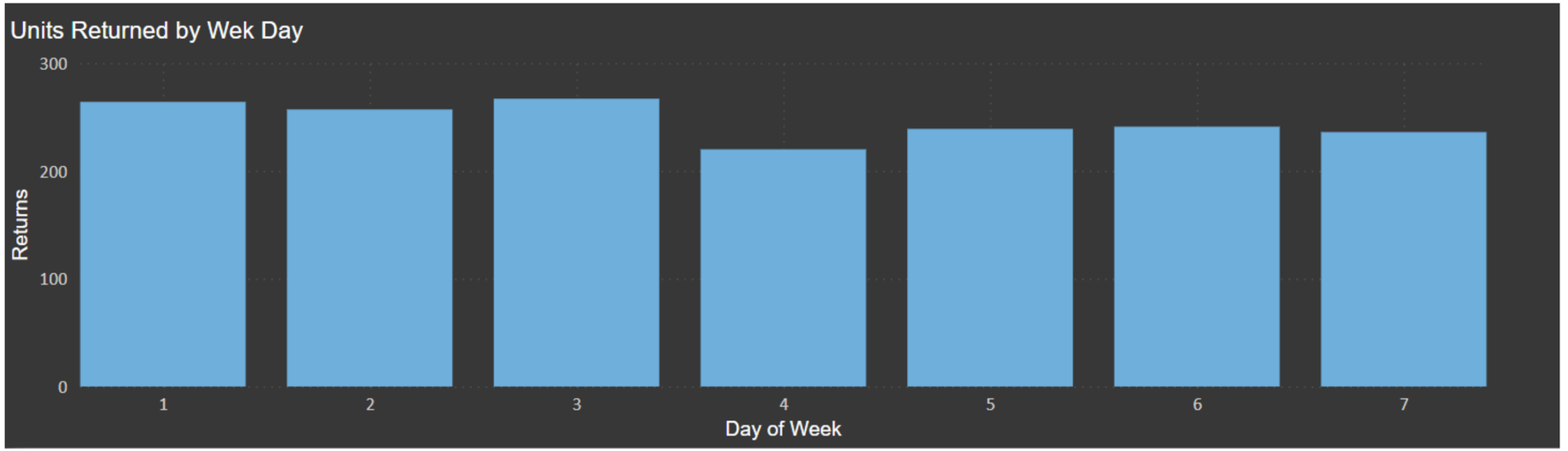
Next Page

Categories of Charts

+ CHANGE OVER TIME

Measures: 1 or more
Dimensions: 1 (Discrete Time)

Bar Chart



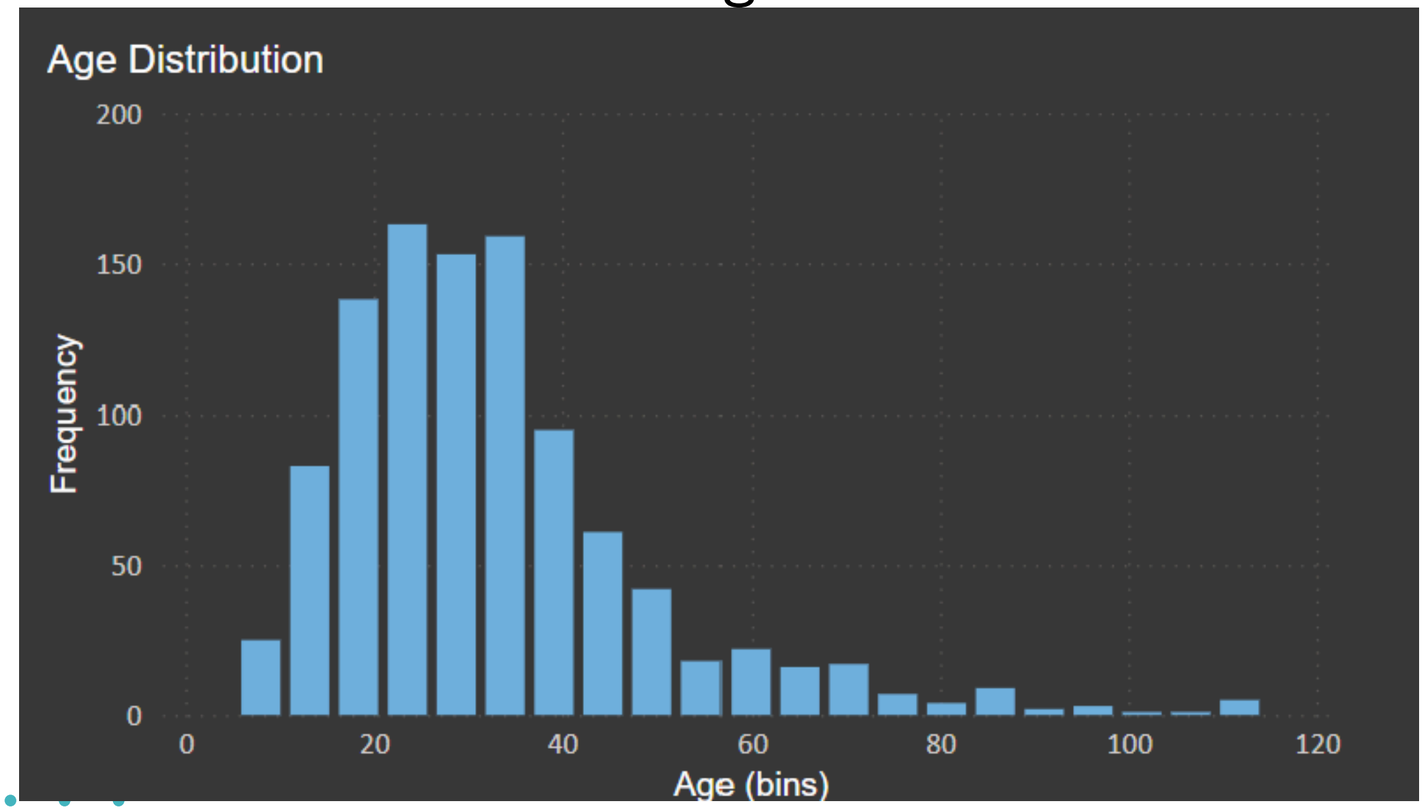
Next Page

Categories of Charts

+ DISTRIBUTIONS

Measures: 1
Dimensions: 0

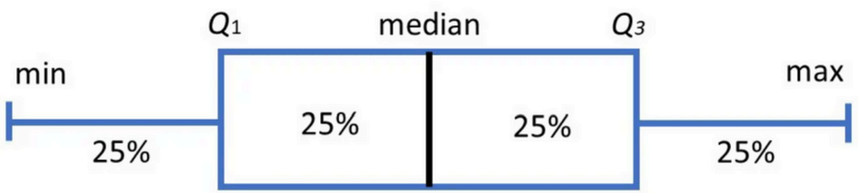
Histogram

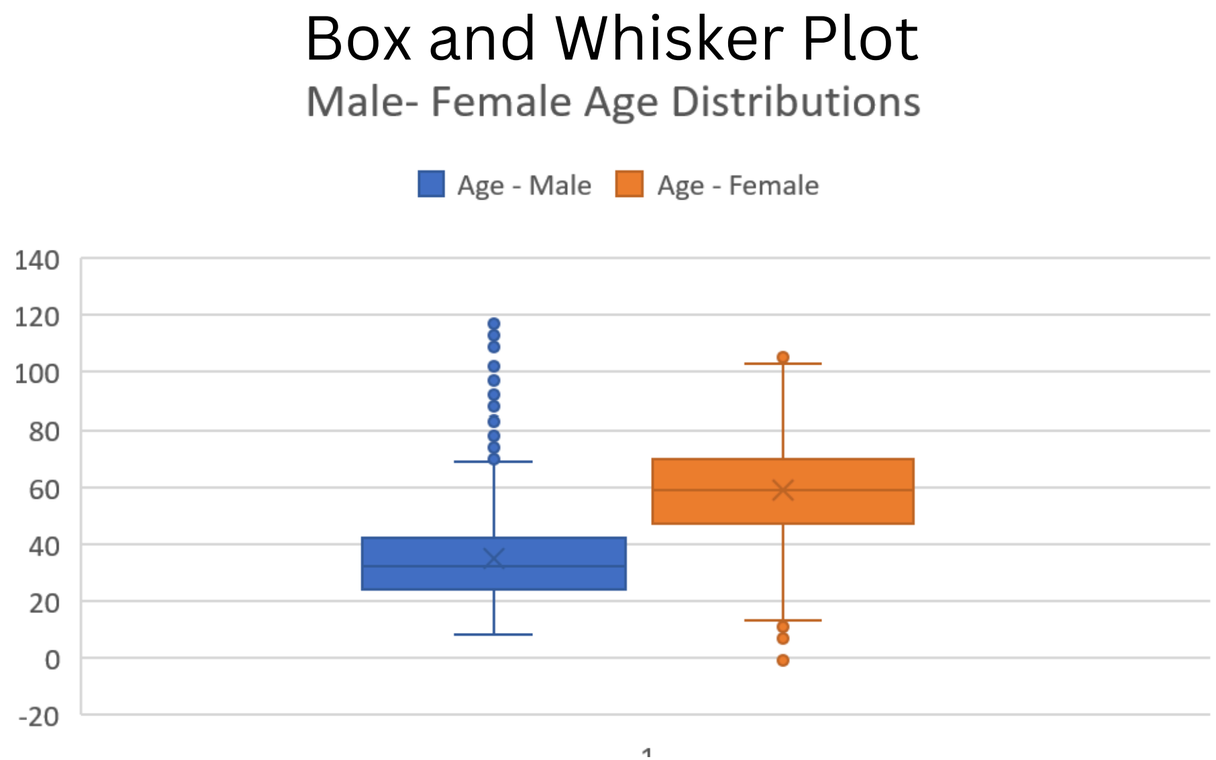


Categories of Charts

+ DISTRIBUTIONS

Measures: 1
Dimensions: 0





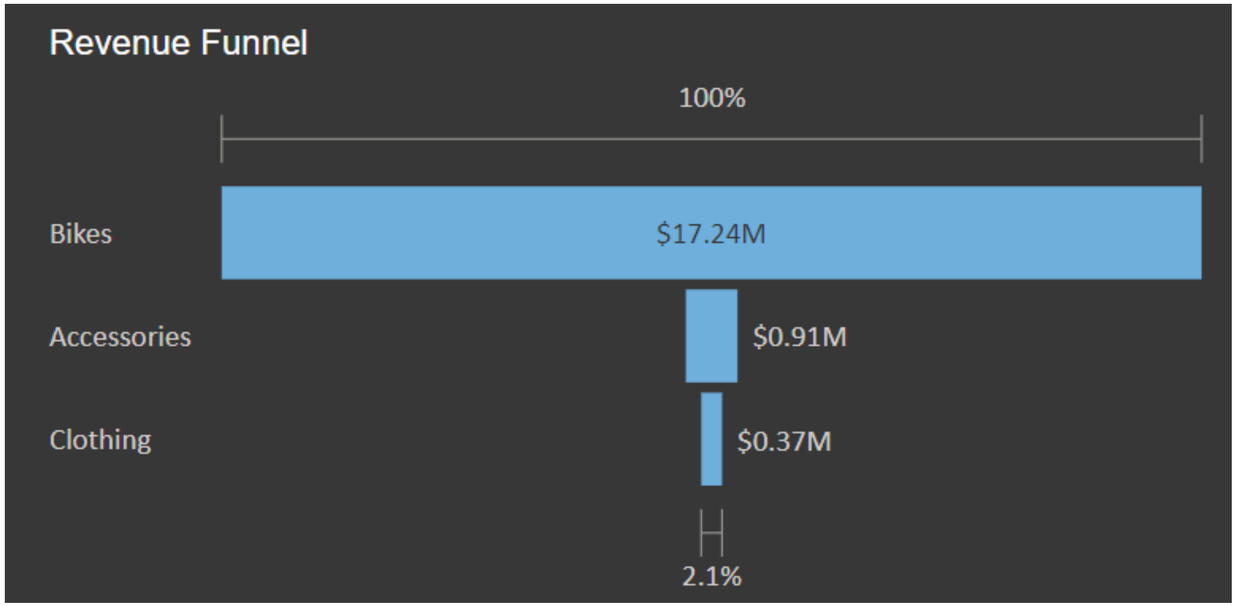
Next Page

Categories of Charts

+ DISTRIBUTIONS

Measures: 1
Dimensions: 1

Funnel Chart



Next Page

Categories of Charts

+ COMPARISONS

Measures: 1
Dimensions: 1 or more

1 - Bar Chart

2 - Stacked Bar Chart

3 - Clustered Bar Chart

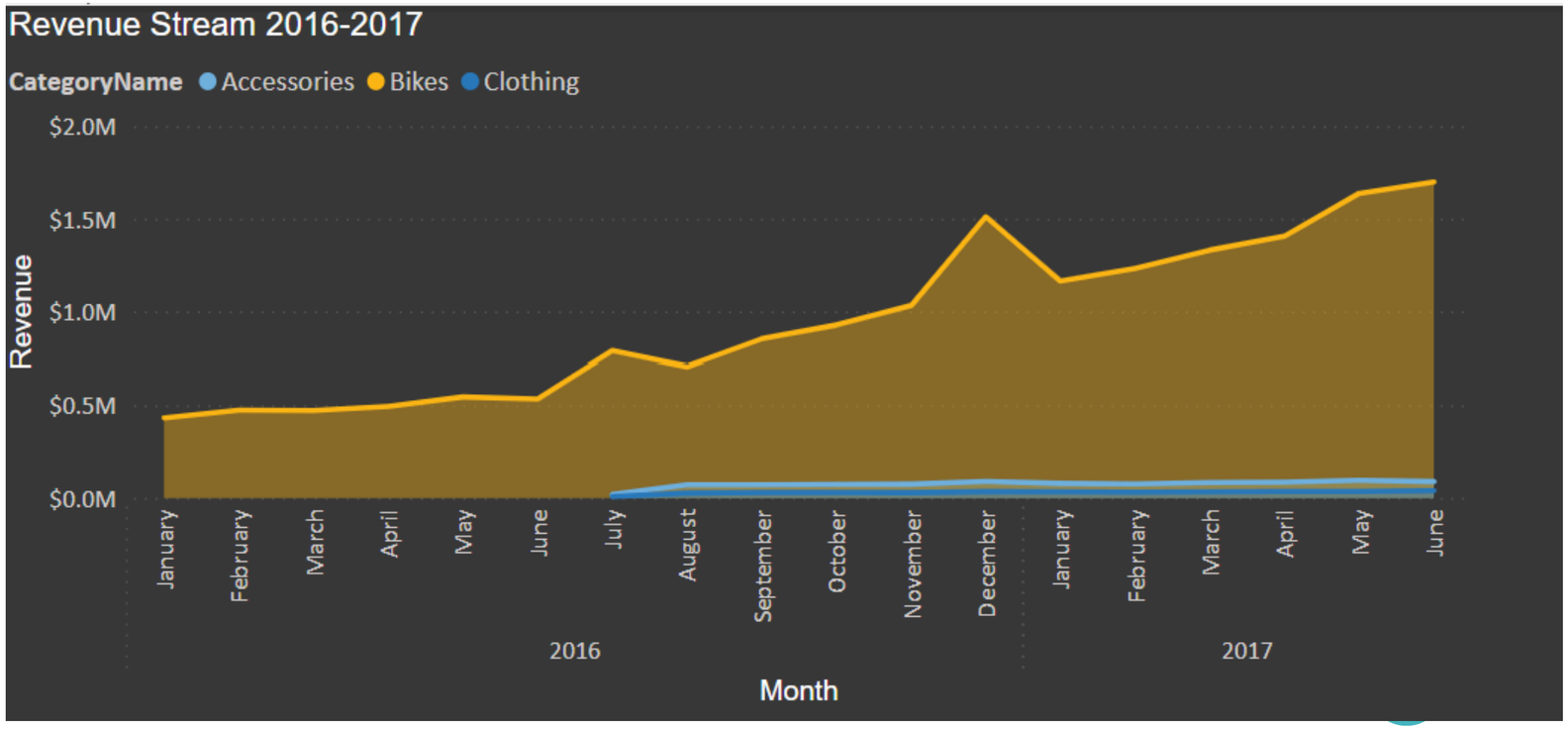
4 - 100% Stacked Bar Chart

Categories of Charts

+ COMPARISONS

Measures: 1
Dimensions: 1 or more

Area Chart

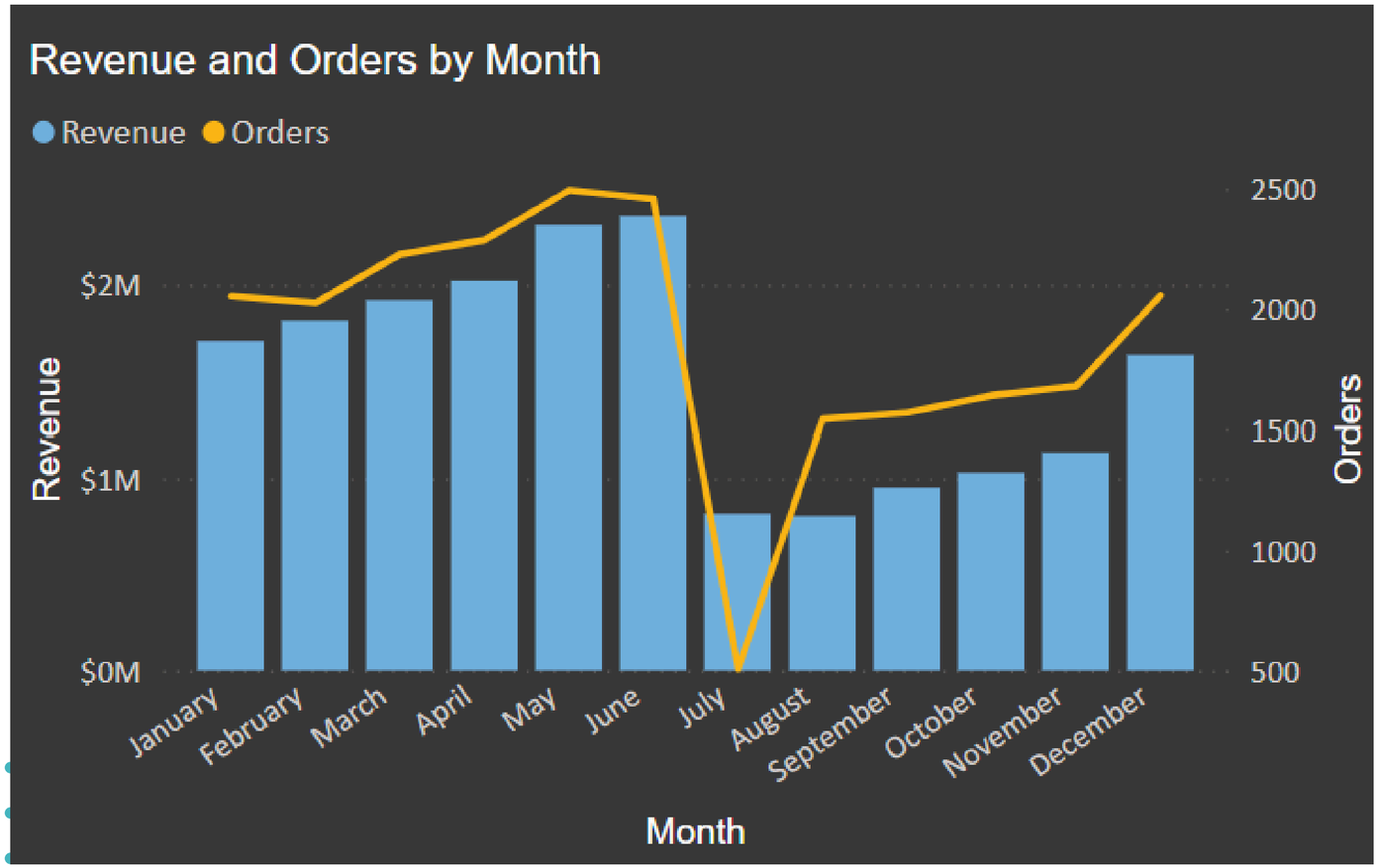


Categories of Charts

+ COMPARISONS

Measures: 2
Dimensions: 1 or more

Combo Chart

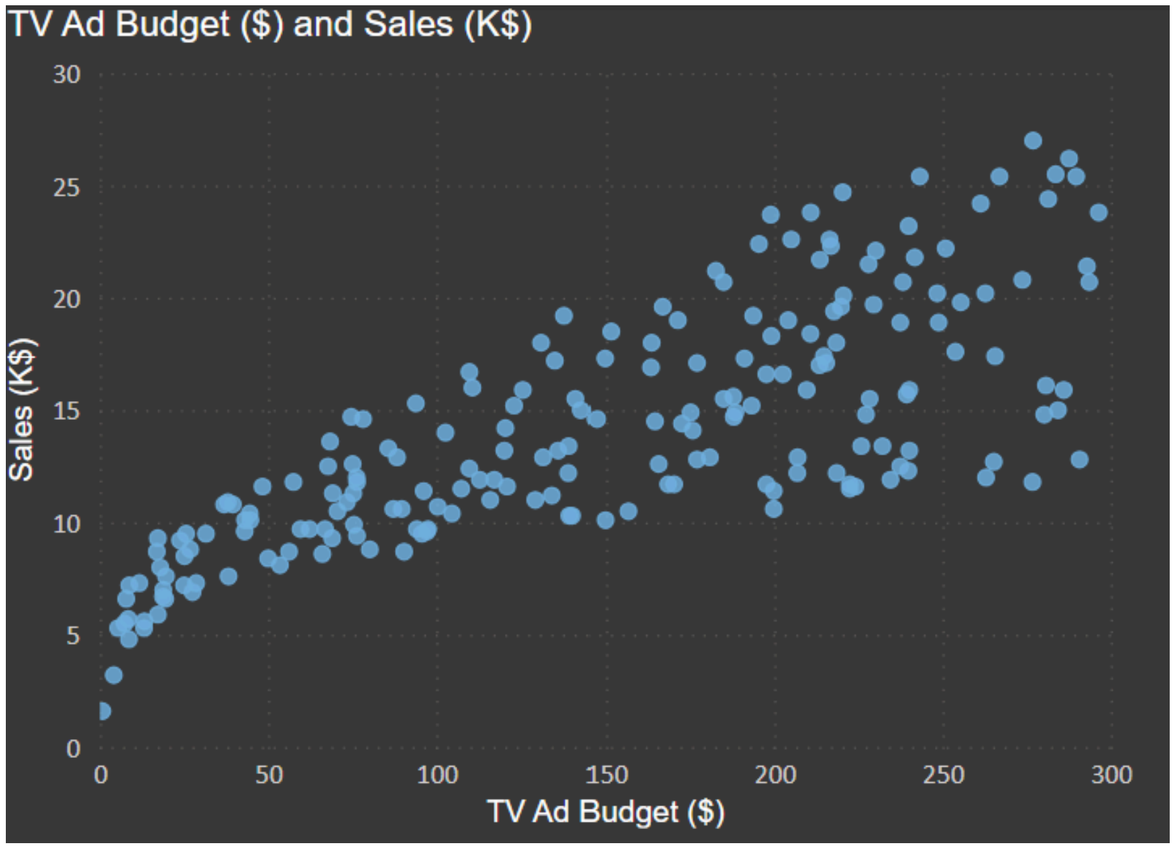


Categories of Charts

+ RELATIONSHIPS

Measures: 2
Dimensions: 0 or more

It represents the correlation between two numeric variables.



Next Page

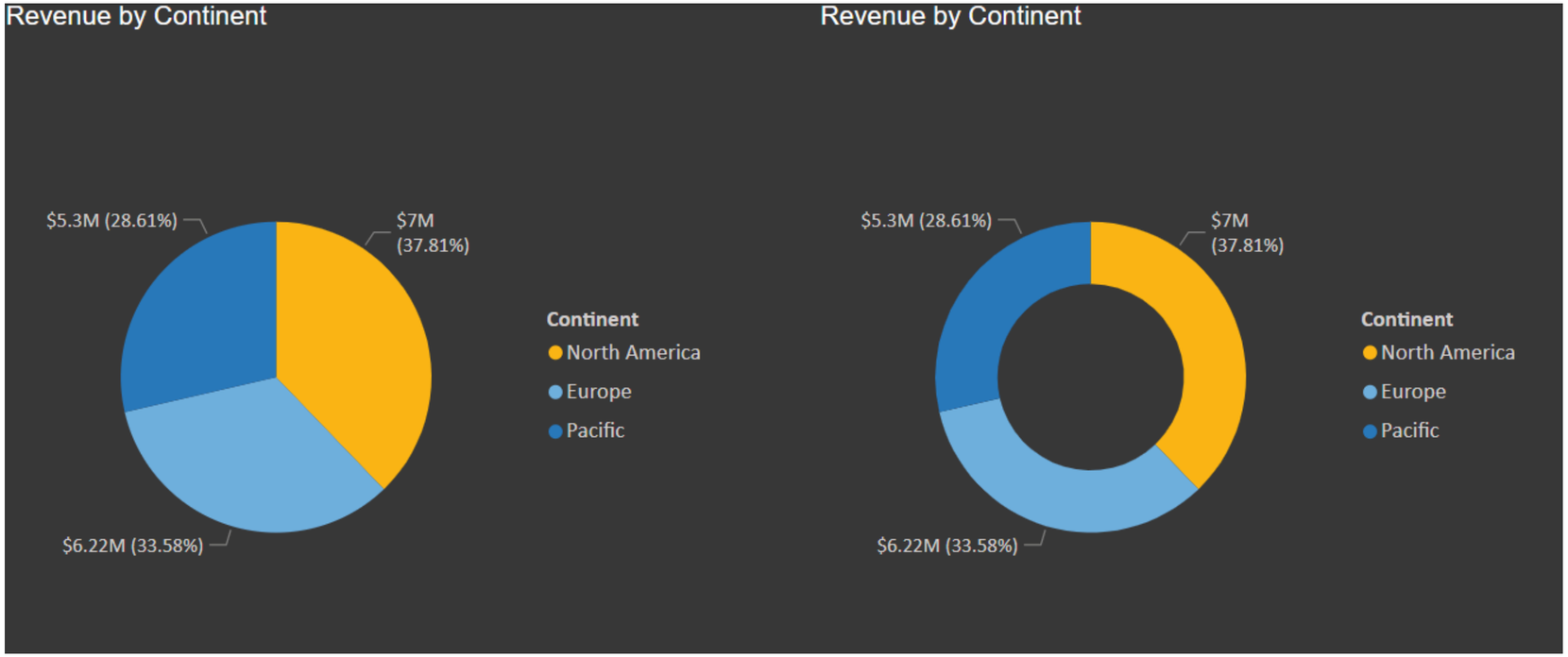
Categories of Charts

+ COMPSITIONS

Measures: 1
Dimensions: 1

Not suitable for more than 5 categories

Pie and Donut Charts



Next Page

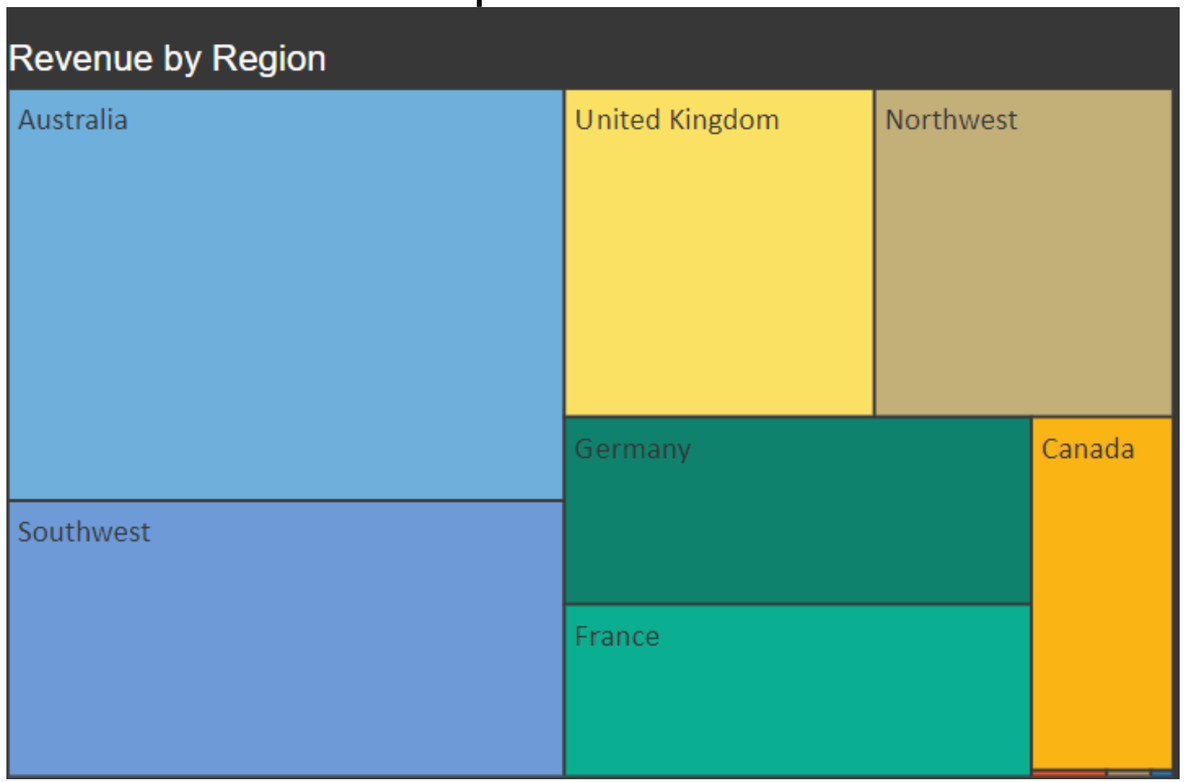
Categories of Charts

+ COMPSITIONS

Measures: 1
Dimensions: 1

suitable for more than 5 categories

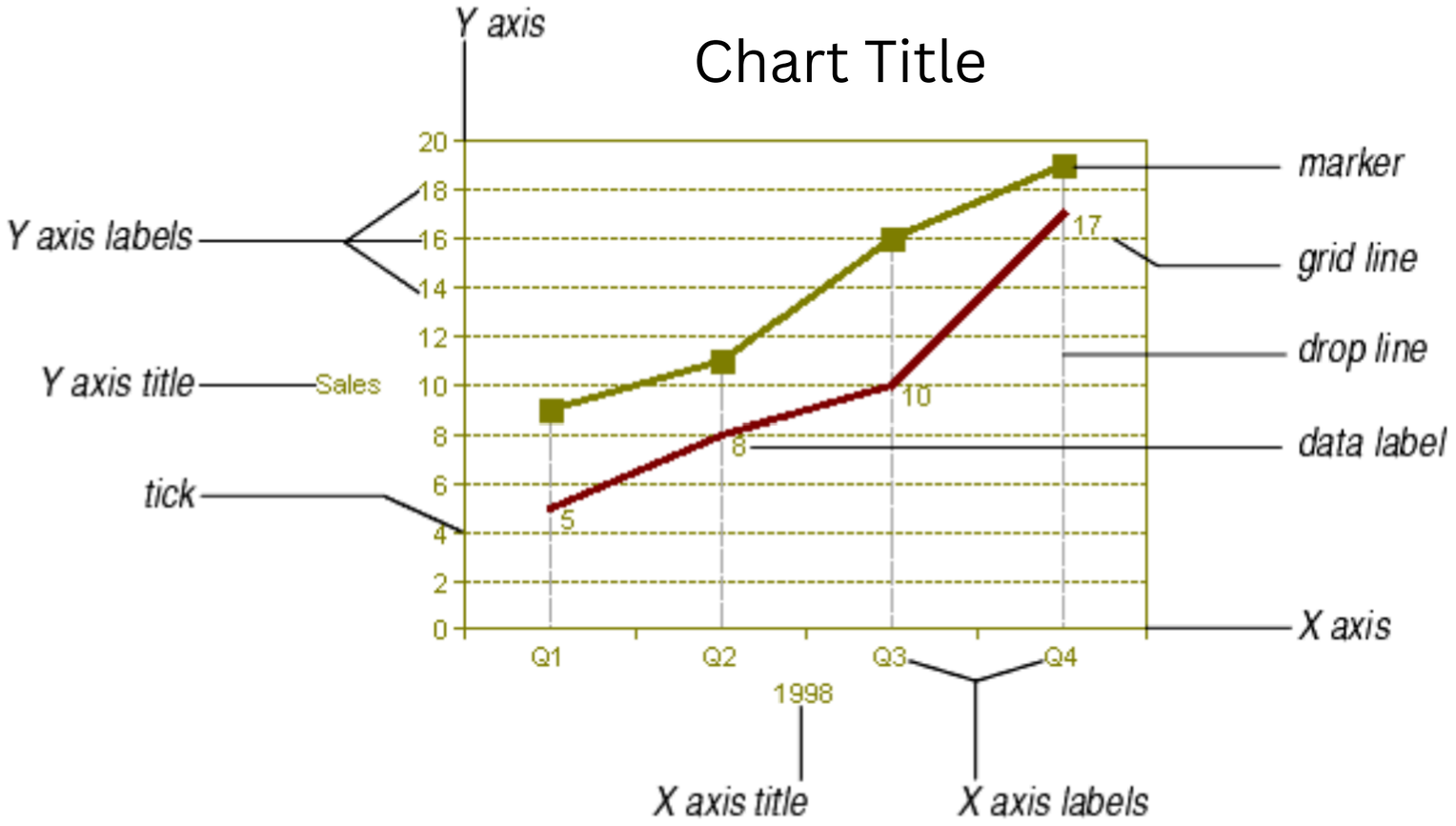
Treemap Chart

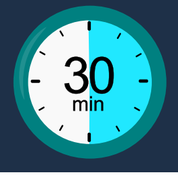


Next Page

Chart Elements

1. Chart Title
2. Grid Lines
3. Drop Lines
4. X axis Label
5. Y axis Label
6. X axis Title
7. Y axis Title
8. Data Labels





Exploratory Data Analysis (EDA)

Next Page

Exploratory Data Analysis

Exploratory Data Analysis (EDA) is a crucial step in the data analysis process. It involves **investigating**, **summarizing**, and **visualizing** data to gain **insights**, **identify patterns**, and formulate **hypotheses** before applying more advanced analytical techniques.

Exploratory Data Analysis - Objectives

- **Gain a general understanding of the data:** Identify the variables, their types (measures, dimensions), and any potential issues like missing values or outliers.
- **Discover patterns and relationships:** Look for trends, correlations, and anomalies within the data.
- **Formulate initial hypotheses:** Based on your observations, develop tentative explanations for the patterns and relationships you discover.

Exploratory Data Analysis - Techniques

- **Descriptive statistics:** Calculate summary measures like mean, median, standard deviation, and frequency distributions to understand the central tendency, spread, and distribution of the data.
- **Data visualization:** Create various charts and graphs like histograms, scatter plots, boxplots, and heatmaps to visually represent the data and identify patterns.
- **Correlation analysis:** Measure the strength and direction of the linear relationship between two variables.

Exploratory Data Analysis - Descriptive Statistics

- **Central Tendency Measure:** Calculate summary measures like mean, median, and mode. Check for anomalies and outliers.
- **Spread Measures:** Calculate summary measures like variance, and standard deviation to measure how is your data is spread around the central value.
- **Check for anomalies and outliers:** Based on the above two measures, we can decide if data has outliers or anomalies.
- **Check for correlations**: examine the relationships between two variables at a time to check if there is any kind of correlation.