

Relatório do Projeto de Power BI



Data Science & Business Analytics - 13.^a Edição
Business Intelligence & Decision Support

Sumário

01 | Enquadramento

02 | Metodologia

03 | Estrutura do Power BI

04 | Análise e Conclusões

05 | Referências

1. Enquadramento

O presente relatório visa suportar o projeto de Power de BI produzido sobre o *dataset* da Adventure Works, uma empresa fictícia do setor do ciclismo. O projeto teve como objetivo identificar e responder a questões críticas de negócio, utilizando o Power BI como ferramenta de análise visualização de dados.

A escolha deste *dataset* prendeu-se com a riqueza dos seus dados e a diversidade de tabelas disponíveis, o que permitiu aplicar os conhecimentos adquiridos em contexto de aula na realização de análises mais abrangentes.

Para além de ser especializada na venda de produtos de ciclismo, a Adventure Works comercializa igualmente outros produtos tais como acessórios e vestuário, tendo operações em várias regiões do mundo como a América do Norte, Austrália e Europa.

A análise da atividade da empresa cingiu-se ao período de janeiro de 2015 até junho de 2017 e procurou responder a cinco questões específicas sobre as vendas, o perfil de clientes e o desempenho dos produtos, a saber:

1. Qual o país com o maior volume de vendas no primeiro semestre de 2017?
2. Qual o continente com melhor desempenho de vendas no segmento de bicicletas?
3. Qual o perfil dos clientes do segmento de vestuário na Europa?
4. Qual é o segmento de produtos mais lucrativo?
5. Qual é o produto com mais devoluções em 2016?

2. Metodologia

A realização do projeto compreendeu as quatro principais fases do processo de análise de dados: importação de dados, preparação de dados, modelação de dados e visualização de dados.

A. Importação de Dados

Esta fase envolveu a conversão do *dataset* e o seu carregamento no Power BI através do editor do Power Query.

O *dataset* original foi extraído do Kaggle e continha 10 ficheiros csv. Por forma a cumprir o requisito de integração de 3 tipos de fontes distintas, alguns dos ficheiros foram convertidos, sendo o *dataset* final composto por:

- 3 ficheiros xlsx
 - AdventureWorks_Customers.xlsx
 - AdventureWorks>Returns.xlsx
 - AdventureWorks_Territories.xlsx
- 3 ficheiros csv
 - AdventureWorks_Product_Categories.csv
 - AdventureWorks_Product_Subcategories.csv
 - AdventureWorks_Products.csv
- 1 pasta com 3 ficheiros csv
 - AdventureWorks_Sales_2015.csv
 - AdventureWorks_Sales_2016.csv
 - AdventureWorks_Sales_2017.csv

Nota: o ficheiro AdventureWorks_Calendar.csv foi desprezado uma vez que foi criada uma tabela de calendário no Power BI com recurso à função DAX 'CALENDAR'.

B. Preparação de Dados

Nesta fase utilizou-se o editor do Power Query para proceder à limpeza e transformação dos dados importados.

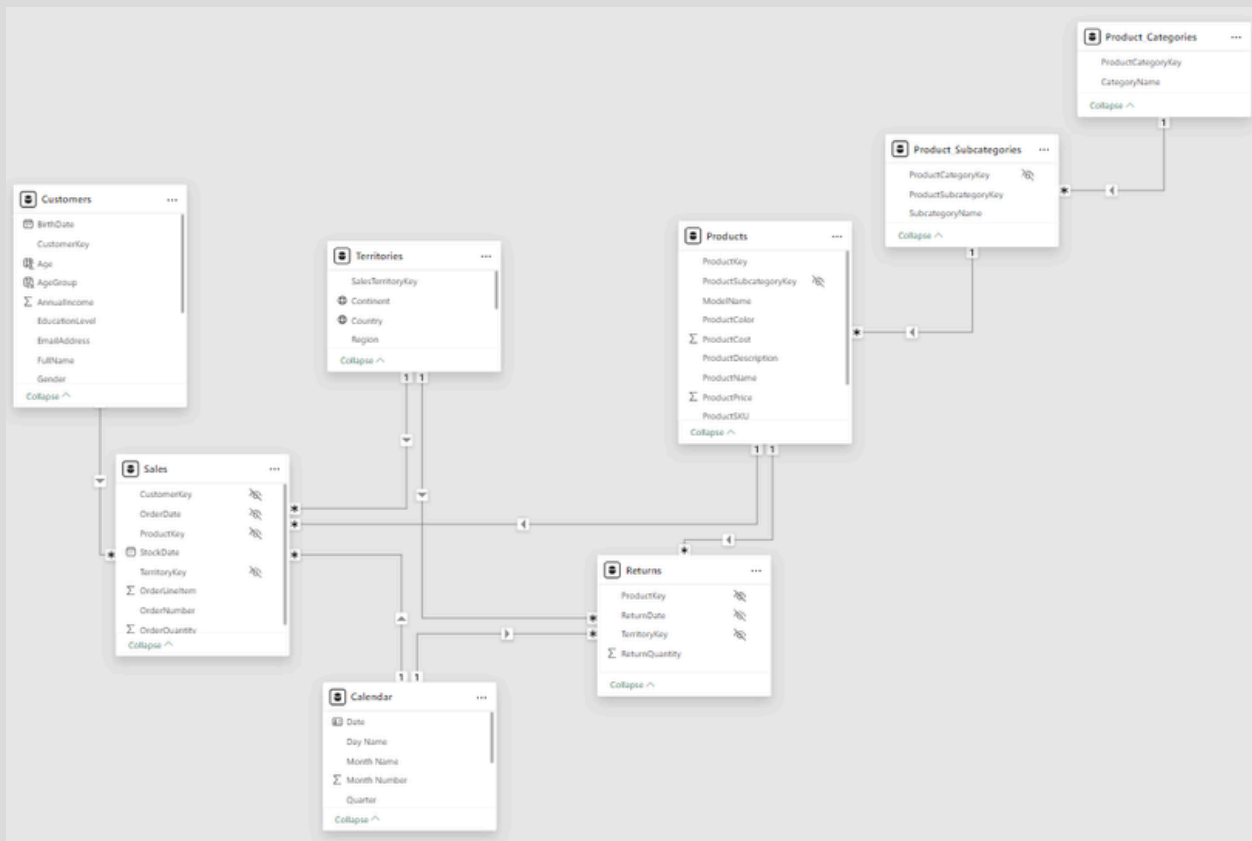
Neste âmbito, foram realizadas transformações tais como: renomeação de tabelas e colunas, promoção de cabeçalho, união de colunas, remoção de colunas e substituição de valores.

Business Intelligence & Decision Support

C. Modelação de Dados

Esta fase envolveu a criação de relações entre as tabelas, bem como a criação de métricas e colunas calculadas.

O modelo de dados é composto por 2 tabelas de factos e 6 tabelas de dimensões, assumindo um *snowflake schema*.



Nota: No modelo de dados as colunas referentes a chaves estrangeiras nas tabelas de factos foram escondidas de modo a garantir que apenas as chaves primárias das tabelas de dimensões eram utilizadas na análise.

Nesta fase foram ainda criados cálculos personalizados (métricas e colunas calculadas), utilizando a linguagem DAX (Data Analysis Expressions), de modo a enriquecer a análise e visualização dos dados.

Exemplos de métricas

- Total Revenue = SUMX(Sales, Sales[OrderQuantity] * RELATED(Products[ProductPrice]))
- Gross Profit = [Total Revenue] - [Total Product Cost]
- Return Rate = DIVIDE([Quantity Returned],[Quantity Sold],"No Sales")
- Total Orders = DISTINCTCOUNT(Sales[OrderNumber])
- YTD Revenue = CALCULATE([Total Revenue], DATESYTD('Calendar'[Date]))

Business Intelligence & Decision Support

Nota: Para fins de organização do modelo de dados todas as métricas foram criadas numa tabela dedicada denominada 'Metrics'.

Exemplos de colunas calculadas (tabela 'Customers')

- `IncomeLevel = SWITCH(TRUE(), [AnnualIncome] >= 150000, "Very High", [AnnualIncome] >= 100000, "High", [AnnualIncome] >= 50000, "Average", "Low")`
- `Age = DATEDIFF(Customers[BirthDate], MAX('Calendar'[Date]), YEAR)`
- `AgeGroup = SWITCH(TRUE(), [Age] >= 60, "Old-age (60+)", [Age] >= 45, "Middle-age (45+)", [Age] >= 25, "Adult (25+)", "Young-age")`

Adicionalmente, foram feitas outras transformações ao nível do Power BI para tornar a experiência do utilizador final mais *"user friendly"*, nomeadamente:

- Criação de hierarquia na tabela 'Territories' composta pelas colunas 'Continent', 'Country' e 'Region';
- Criação de um grupo de dados na tabela 'Customers' sobre a coluna 'Occupation';
- Criação de sinónimos (e.g. total revenue, revenue, sales).

D. Visualização de Dados

Esta fase envolveu a construção de um relatório interativo utilizando gráficos, tabelas e outros elementos visuais para apresentar os *insights* de forma clara e intuitiva.

A mesma é detalhada no capítulo seguinte onde é apresentada a estrutura do relatório Power BI, incluindo a identificação das análises e visuais utilizados.

3. Estrutura do Power BI

O Power BI compreende 3 páginas distintas que oferecem uma análise integrada do negócio da Adventure Works: desempenho de vendas, características dos clientes e desempenho dos produtos.

1. Sales Performance Overview - Oferece uma visão geral do desempenho de vendas da empresa, permitindo uma análise detalhada das vendas por região, período temporal e produtos.

- Total de vendas e pedidos (*Cards*);
- Top 10 Selling Products (*Matrix*);
- Tendência de vendas por continente (*Line chart*);
- Total de vendas, pedidos e lucro bruto (*Map*);
- Vendas por subcategoria (*Stacked bar chart*);
- Vendas por país (*Clustered column chart*).

2. Customer Insights - Apresenta o perfil de cliente da empresa segundo as variáveis rendimento, género, idade e ocupação, analisadas pelo total de vendas.

- Total de vendas e número de clientes (*Cards*);
- Vendas por género e rendimento (*Donut chart*);
- Vendas por ocupação (*Stacked column chart*);
- Vendas por (*Clustered bar chart*);
- Cliente por total de vendas, pedidos e valor médio por pedido (*Matrix*).

3. Product Performance Analysis - Inclui uma análise do desempenho do portefólio de produtos da empresa como, por exemplo, número total de produtos, categorias de produtos mais lucrativas, margem bruta e taxa de devoluções.

- Total de categorias, subcategorias e de produtos (*Cards*);
- Total de vendas, pedidos e margem bruta por categoria (*Bubble / Scatter Chart - AppSource visual*);
- Vendas por categorias e subcategoria (*Treemap*);
- Produto por margem bruta e taxa devolução (*Matrix*).

As três páginas do Power BI incluem, igualmente, 3 *slicers* que permitem a filtragem dos dados por Continente, Categoria de Produto e Data. Os *slicers* estão sincronizados entre todas as páginas, garantindo uma análise consistente por parte do utilizador final.

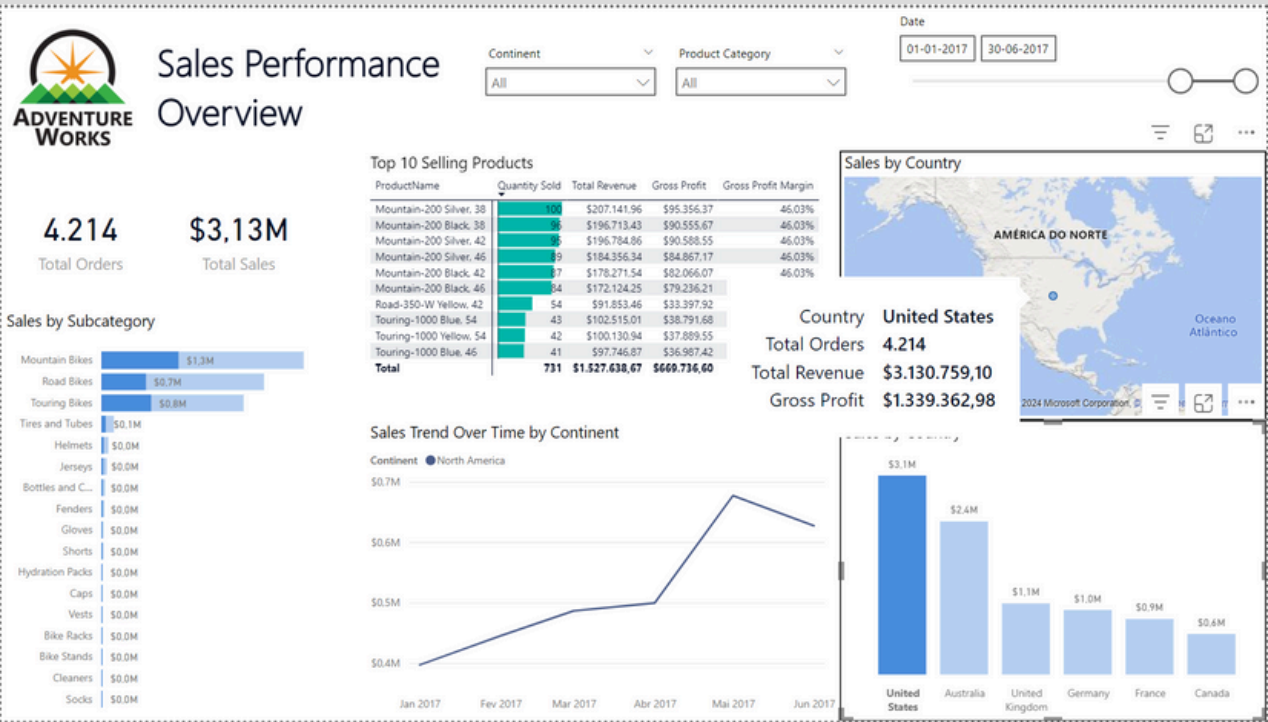
4. Análise e Conclusões

Neste capítulo apresentamos uma análise detalhada das questões elencadas no capítulo Enquadramento, focando em *insights* significativos que emergiram da análise e exploração dos dados.

1. Sales Performance Overview

01 | Qual foi o país com maior volume de vendas ?

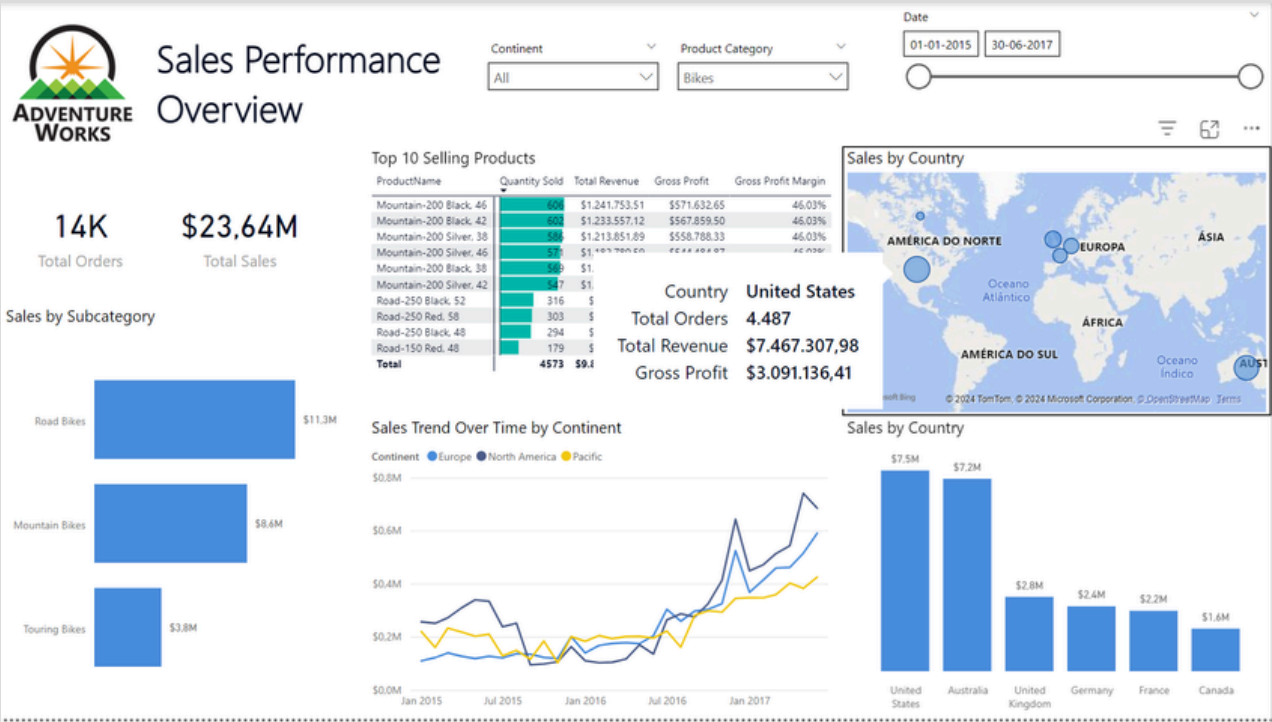
O país com o maior volume de vendas no primeiro semestre de 2017 é os EUA.



4. Análise e Conclusões

02 | Qual o continente com melhor desempenho de vendas no segmento de bicicletas?

A América do Norte é o continente com melhor desempenho de vendas ao longo do tempo no segmento de 'Bikes'.

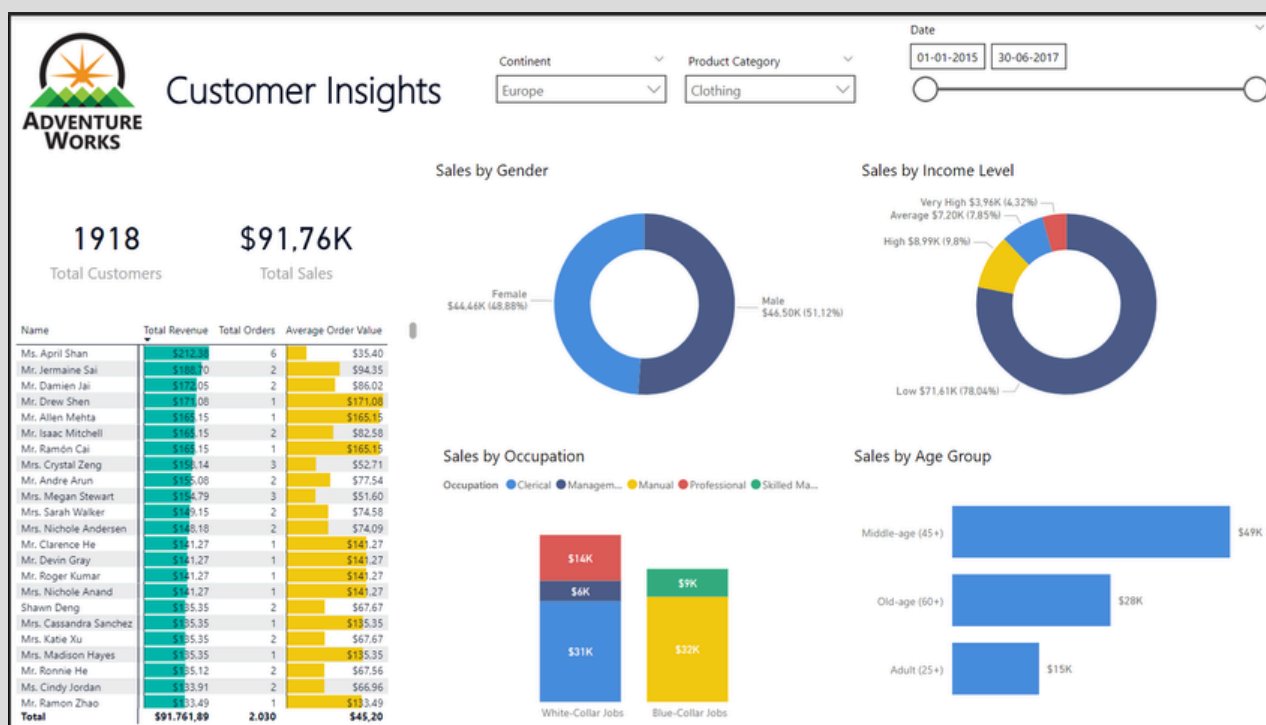


4. Análise e Conclusões

2. Customer Insights

03 | Qual o perfil dos clientes do segmento de vestuário na Europa?

O perfil cliente que gera mais vendas no segmento de 'Clothing' na Europa tem meia idade (45-59), rendimentos baixos (<50K) e emprego em trabalhos manuais ou administrativos.

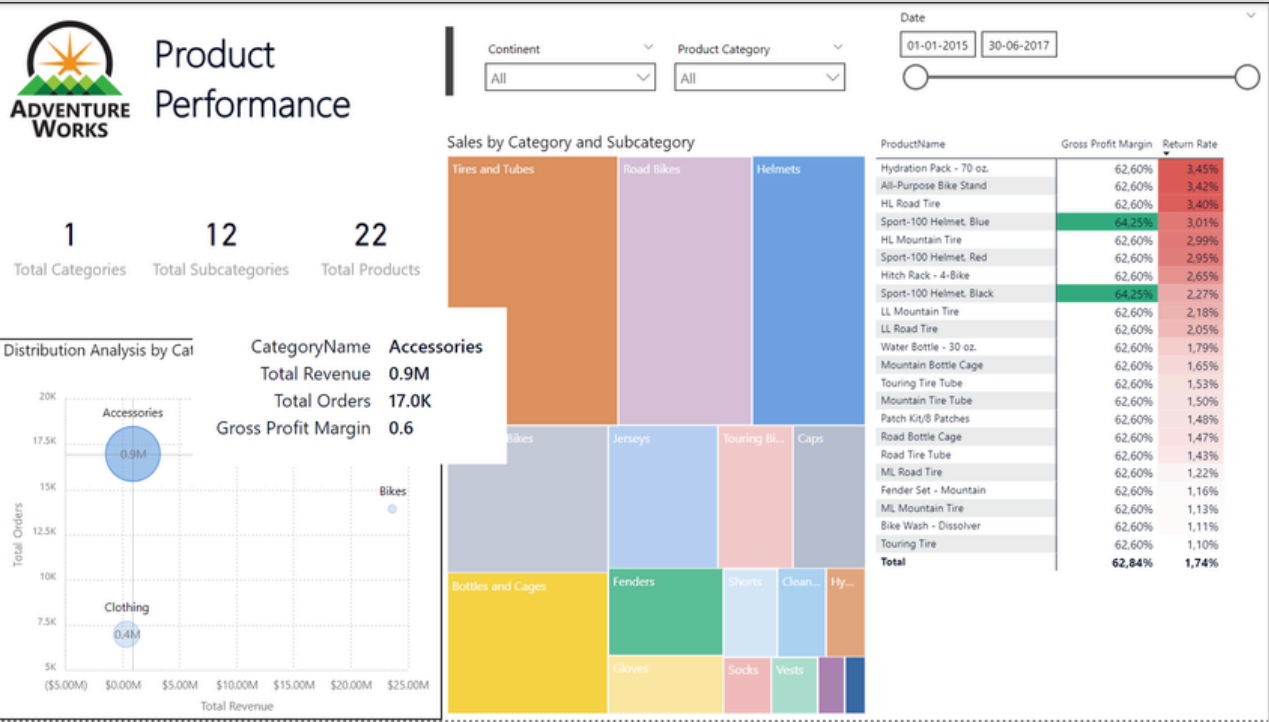


4. Análise e Conclusões

3. Product Performance Analysis

04 | Qual é o segmento de produtos mais lucrativo?

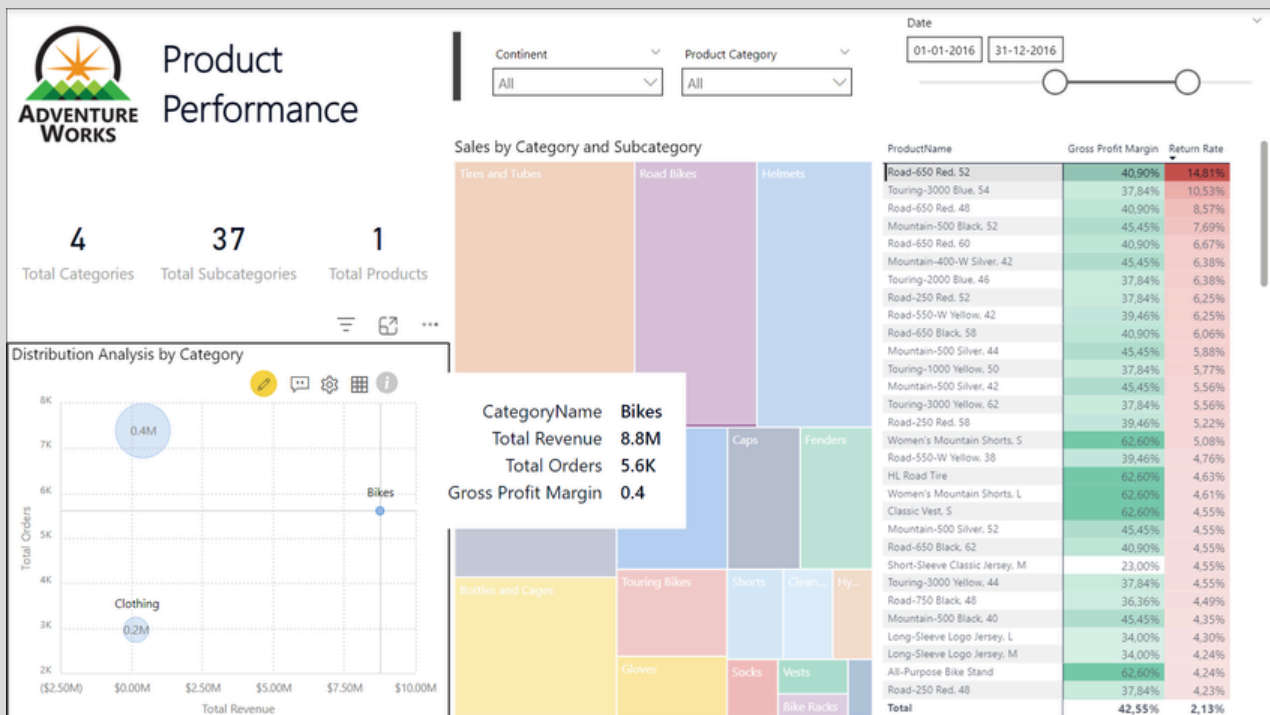
O segmento de produtos mais lucrativo é o 'Accessories' com uma taxa bruta média de ~60%.



Business Intelligence & Decision Support

05| Qual é o produto com mais devoluções em 2016?

O produto com mais devoluções em 2016 foi 'Road 650 Red, 52'.



5. Referências

Kaggle: <https://www.kaggle.com/datasets/ukveteran/adventure-works>

Microsoft AppSource: <https://appsource.microsoft.com/en-us/product/power-bi-visuals/WA200002313?tab=Overview>

Power BI Service: <https://app.powerbi.com/groups/6b12d29e-4224-4b50-bb64-30c665aef3e3/reports/d494787e-cbc0-44d0-baf3-37fbc19c93a7/0dc2a81e94410062d086?experience=power-bi>