KIN+CARTA

Data scientist analytical challenge: “Nike By You” sales forecast divergence analysis

Alberto M. Palacio Bastos

23 February 2024

Presentation by

Alberto Palacio
M.Sc. Data Scientist
alberto.palaciob@gmail.com

Summary

Data enthusiast with **10+ years of experience** in engineering and data projects for the construction, mining, energy, and oil & gas industries.

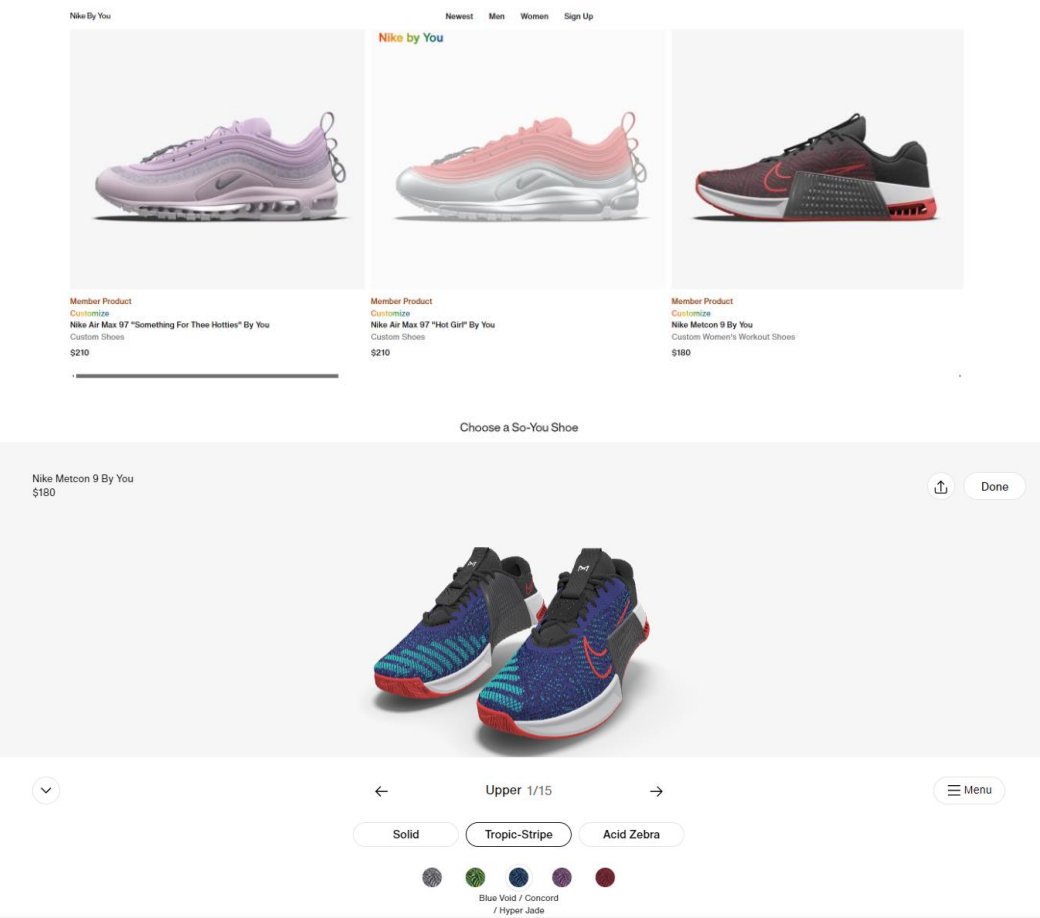
Proficient in data analysis and extracting insights applying the **CRISP-DM** (Cross-Industry Standard Process for Data Mining) and **EDA** (Exploratory Data Analysis) methodologies for intelligent data driven decision making.

With my knowledge in advanced statistical algorithms, **machine learning** and forecasting, I strive to bring innovative, highly efficient, and high-quality technical solutions to businesses.

linkedin.com/albertompalaciobastos github.com/albertompalaciobastos

Exercise Overview

Nike By You



Problem

For decades, Nike has operated with a wholesale **retail-first** model.

Nike By You is the company's new **direct-to-consumer** sales initiative.

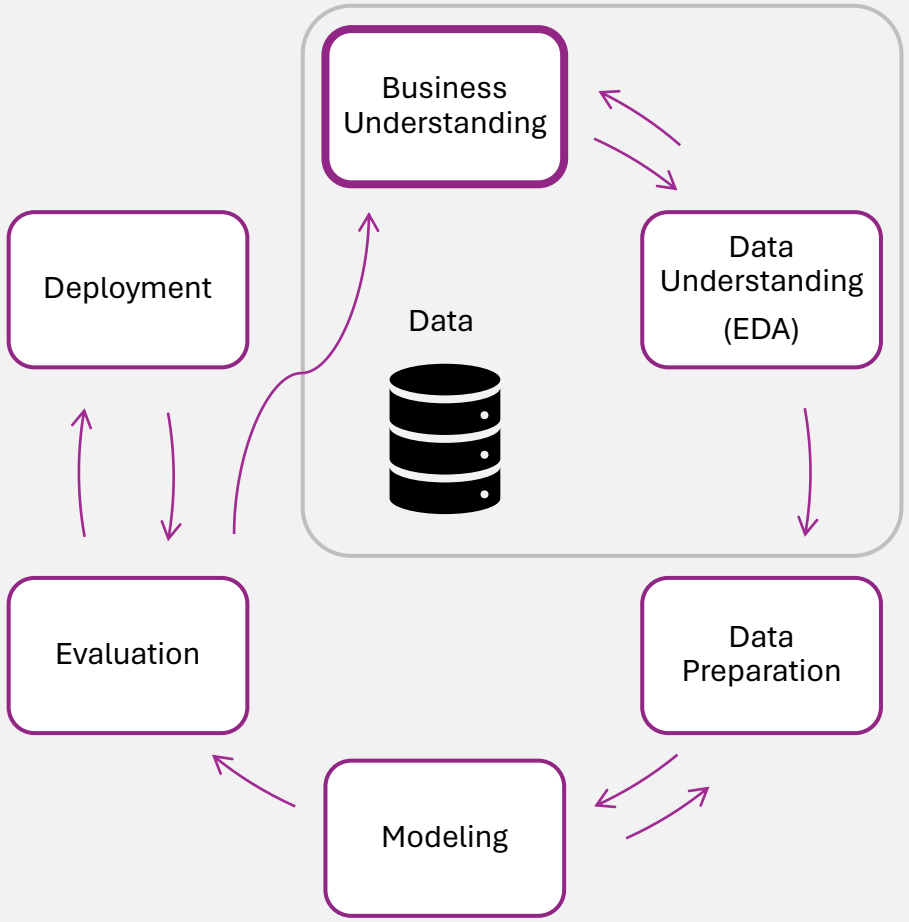
It gives you the chance to **customize** your shoes to your personal taste and color preferences.

The Nike By You business has **missed sales** forecast targets in the past 2 fiscal months.

What **actionable recommendations** would you propose to senior leadership to help inform their decisions on how to improve the Nike By You business?

Analysis Attack Plan

CRISP-DM



1. Business Understanding.

1.1 Understand the question and business needs.

- What happened?
  Missed sales forecast targets.

- Why did this happen?
  They don't know why (**business need**).

- What will happen?
  Forecast targets will still be missed.

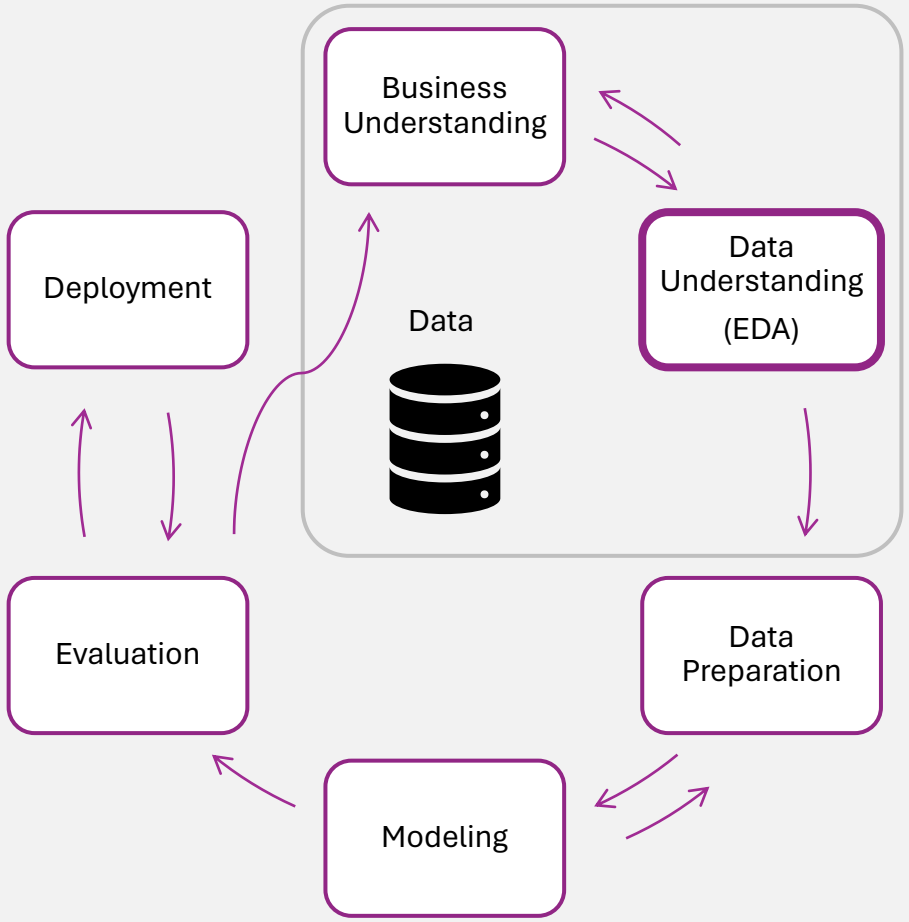
- What should we do?
  Senior executives want **actionable recommendations**.

1.2 Determine appropriate analytic approach.

- Descriptive: Current Status.
- **Diagnostic: Statistical Analysis.** **Root Cause Analysis.**
- Predictive: Forecasting.
- Prescriptive: Recommendations.

Analysis Attack Plan

CRISP-DM



2. Data Understanding.

2.1 Data requirements.

- Asses initial data collection
- Availability
- Quality
- Content

2.2 Data collection.

- Sources
- Elements
- Acquisition Strategies
- Decisions on unavailable data

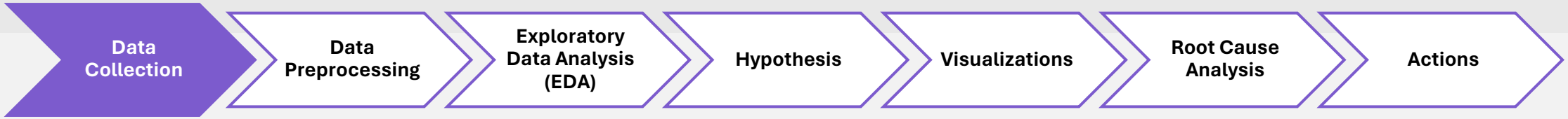
2.3 Data Integration.

- Extraction
- Merge data
- Assessing duplicates and missing data

2.4 Exploratory Data Analysis (EDA)

- Data wrangling
- Data visualizations
- Statistical analysis
- Hypothesis testing

Diagnostic analytics approach: Root cause analysis



- **Internal data:**
  - Website traffic
  - Website clicks
  - Sales/Orders history
  - Shipping history
  - Product metadata
  - Customer data
  - Marketing expenditure
  - SEO system data
  - Customer satisfaction survey

- **External data:**
  - Global economic indicators
  - Historic market size
  - Historic market share

Note: Consider forecast model development **methodology** and **input variables** (internal and external).

Diagnostic analytics approach: Root cause analysis

- **Data cleaning**
  - Remove duplicates
  - Assess missing values
  - Remove unnecessary fields
  - Remove customer personal information (if present)
  - Time series grouping (example: daily sales)

- **Feature Engineering**
  - Define and calculate KPIs and metrics.
    - Abandoned carts
    - Failed purchase attempts
    - Average navigation time
    - Website conversion rate
    - Return on ad spend (ROAS)
  - Dimension reduction strategies

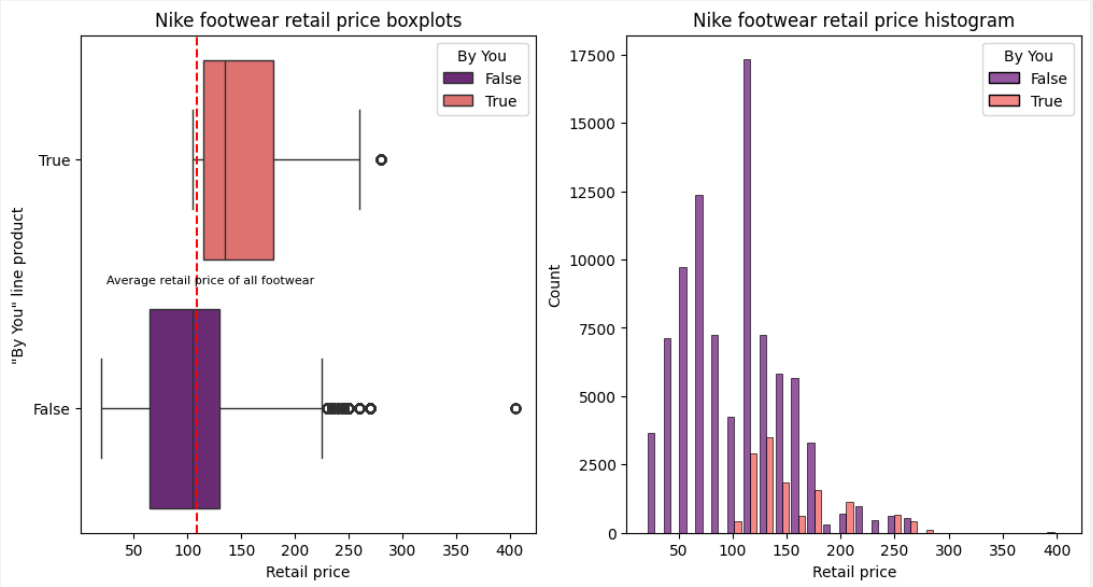
Diagnostic analytics approach: Root cause analysis

Data Collection → Data Preprocessing → **Exploratory Data Analysis (EDA)** → Hypothesis → Visualizations → Root Cause Analysis → Actions

Descriptive statistics

	DEPARTMENT	CATEGORY	SUBCATEGORY	SKU	SKU_VARIANT	PRODUCT_NAME	PRODUCT_ID
count	229471	229471	229471	2.294710e+05	2.294710e+05	229471	2.294710e+05
unique	3	3	29	NaN	NaN	1972	NaN
top	Men	Clothing	All Clothing	NaN	NaN	Nike Sportswear	NaN
freq	118480	126535	52945	NaN	NaN	10604	NaN
mean	NaN	NaN	NaN	4.019402e+07	2.722394e+07	NaN	4.035067e+07
std	NaN	NaN	NaN	1.599420e+08	1.618548e+06	NaN	1.603994e+08
min	NaN	NaN	NaN	1.000072e+07	1.000702e+07	NaN	1.000072e+07
25%	NaN	NaN	NaN	1.369380e+07	2.706616e+07	NaN	1.366746e+07
50%	NaN	NaN	NaN	1.390334e+07	2.765702e+07	NaN	1.390414e+07
75%	NaN	NaN	NaN	1.404521e+07	2.810834e+07	NaN	1.405392e+07
max	NaN	NaN	NaN	1.010262e+09	2.935828e+07	NaN	1.010261e+09

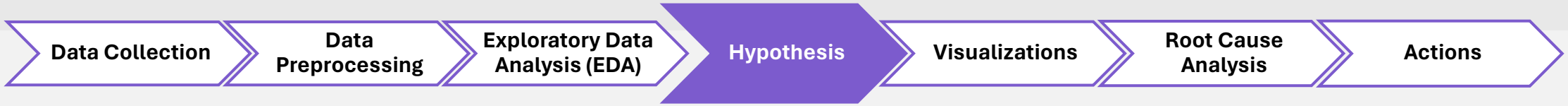
Histograms and boxplots



https://www.kaggle.com/datasets/polartech/nike-sportwear-product-dataset?resource=download

https://github.com/AlbertoMPalacioBastos/nike-by-you-products-analysis/blob/main/Nike_product_data_analysis.ipynb

Diagnostic analytics approach: Root cause analysis



Internal factors/variables

- Low marketing expenditure.
- Wrong marketing strategy.
- Low web-page performance.
- Low product quality or product related issues.
- Manufacturing related issues.
- Shipment/delivery issues.
- Low Customer satisfaction level.

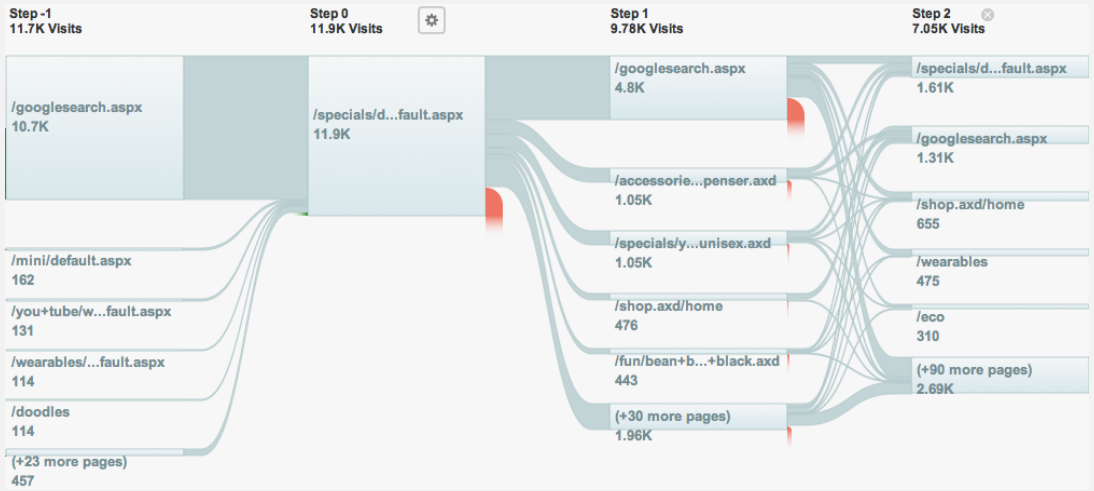
External factor/variables

- External market conditions that the company can not control.
- Competitors launching a similar product.
- Global economy factors.

Diagnostic analytics approach: Root cause analysis

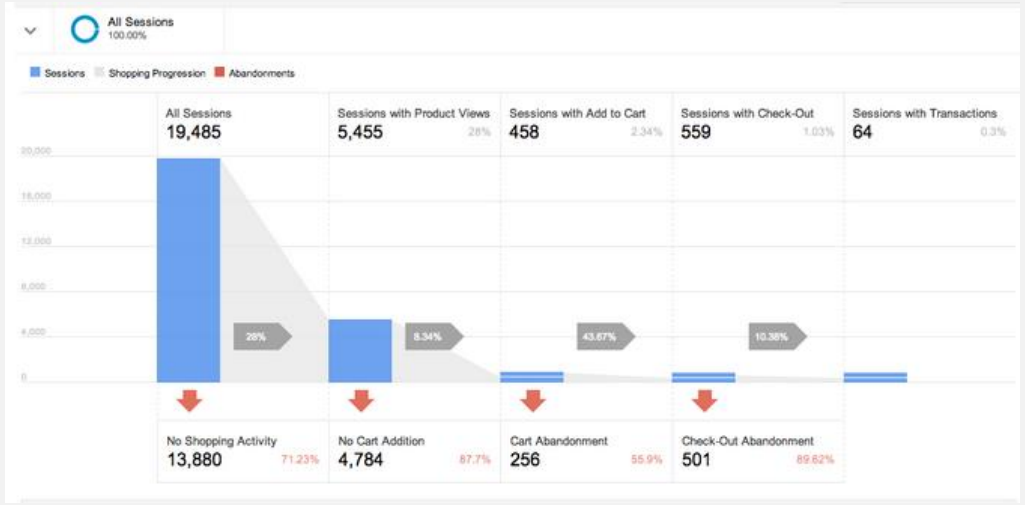
Data Collection | Data Preprocessing | Exploratory Data Analysis (EDA) | Hypothesis | **Visualizations** | Root Cause Analysis | Actions

Visitors' navigation flow



https://analytics.googleblog.com/2011/10/introducing-flow-visualization.html

Sales/goal funnel



https://analytics.googleblog.com/2014/05/better-data-better-decisions-enhanced.html

Diagnostic analytics approach: Root cause analysis

Data Collection | Data Preprocessing | Exploratory Data Analysis (EDA) | Hypothesis | Visualizations | **Root Cause Analysis** | Actions

Clustering machine learning model for customer segmentation

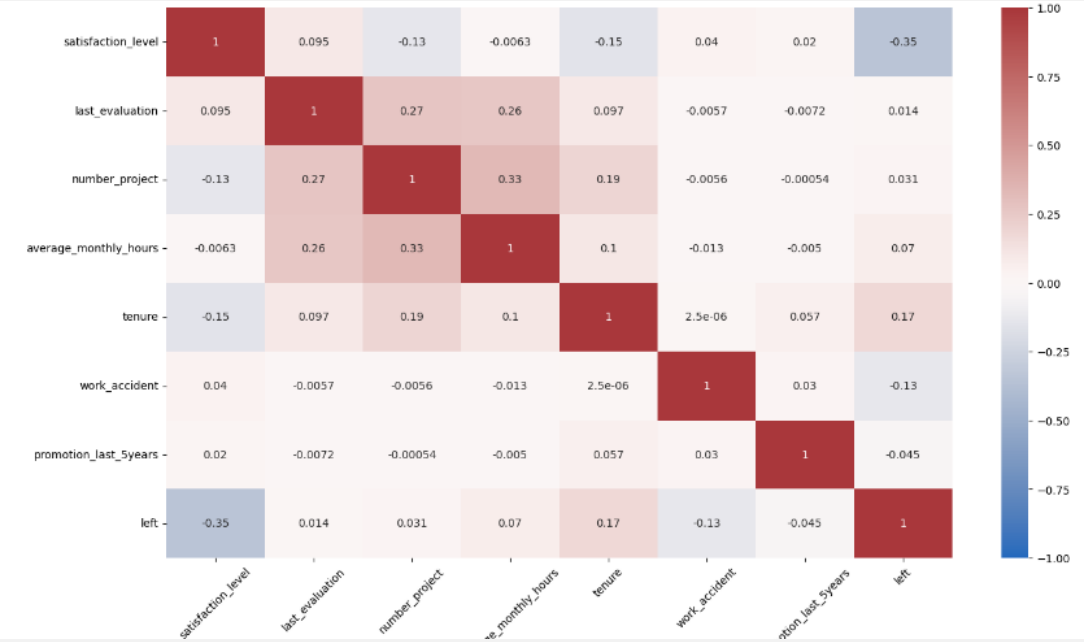
https://medium.com/@robertb909/k-means-clustering-a64f859a1074

- Allows the identification of hidden patterns in the consumer population.
- It would help identify if the marketing strategy is aimed to the right market segment.

Diagnostic analytics approach: Root cause analysis

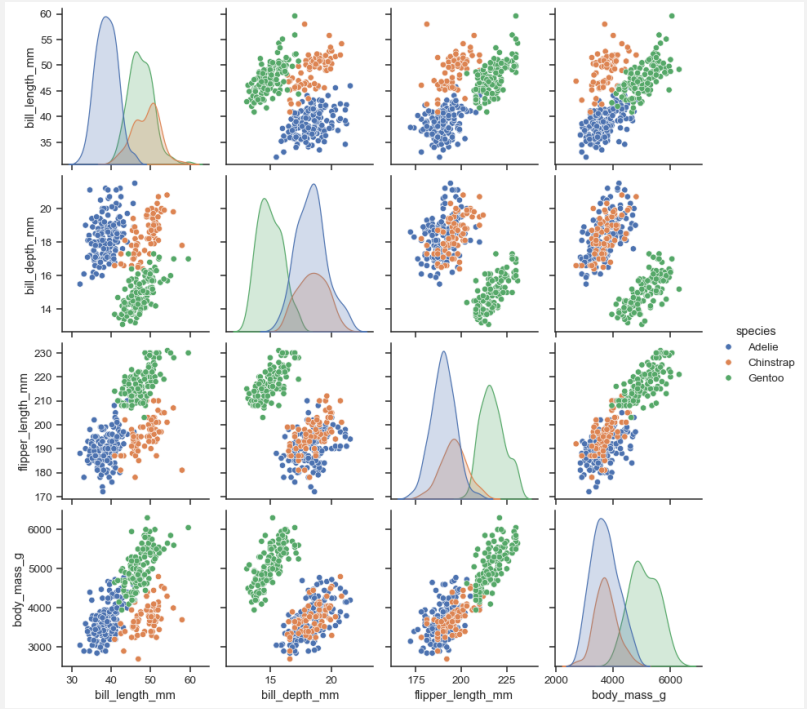
Data Collection | Data Preprocessing | Exploratory Data Analysis (EDA) | Hypothesis | Visualizations | **Root Cause Analysis** | Actions

Correlation heatmaps



https://github.com/AlbertoMPalacioBastos/Machine_Learning_to_provide_insights_for_Human_Resources/blob/main/Salifort%20Motors%20Project.ipynb

Pairwise plots

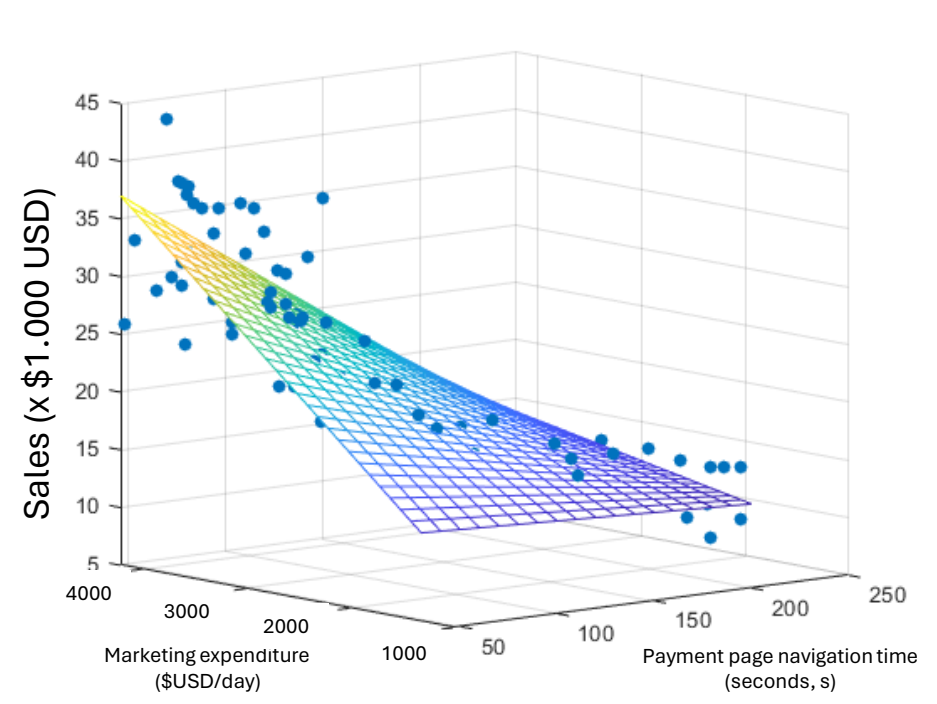


https://seaborn.pydata.org/generated/seaborn.pairplot.html

Diagnostic analytics approach: Root cause analysis

Data Collection | Data Preprocessing | Exploratory Data Analysis (EDA) | Hypothesis | Visualizations | **Root Cause Analysis** | Actions

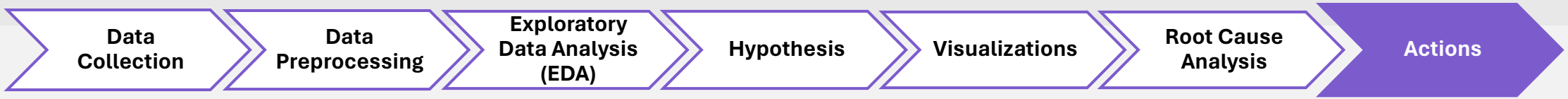
Construct a sales regression model



https://medium.com/analytics-vidhya/new-aspects-to-consider-while-moving-from-simple-linear-regression-to-multiple-linear-regression-dad06b3449ff

- Single or multiple, depending on number of variables with high magnitude coefficients in the correlations heatmap.
- Linear or non-linear, depending on goodness of fit (R-squared coefficient), and considering underfitting and overfitting.
- Allows the identification of variables that have a high impact on sales performance, depending on coefficients t statistic and P-value.
- It would help for forecasting the output of future A/B testing or factorial experiment design, depending on regression coefficients values.

Diagnostic analytics approach: Root cause analysis



Depending on the number of factors or variables identified as the root cause:

Single factor / variable:

- Perform **A/B testing.**
  - Two samples comparison.
  - Means difference statistical t-test.

Multiple factors / variables:

- Perform a **factorial experiment** design and analysis.
  - Two or multiple levels, depending on time and budget restrictions.
  - Perform output optimization by controlling input variables' levels.

Thank You!

Any questions?

Observations and comments:

This exercise assumes that the sales forecasting model is well developed and validated. In a real case scenario, I would evaluate the forecast model, check for errors, inaccurate model assumptions, and risk management strategy also.

KIN+CARTA