* Week 1
  + Name of the project : HCIP (House Construction Investment Project)
  + Objective of the project:
    - Develop a sales model to predict sales by product family
    - Milestones
      * Use new houses dataset started to build by province
      * Establish a correlation with the new construction and Nedco sales dataset
      * Use # of permits by construction type dataset
      * Use Nedco CRM sales data to look into customer profile
      * Evaluate the correlation among sales 2017 / Product family / market segment
      * Evaluate the correlation among sales 2017 / Market segment / ship type
      * Evaluate the correlation among sales 2017 / market segment / ship type / product family
      * Repeat the model with 2016
  + Create a functional map
    - Understand the relationship between:
      * An electrical Distributor (Nedco) & customer
    - Potential attributes from a distributor perspective
      * Sales 2017, product families, market segment, customer type, age of account, profit by customer
      * From a customer perspective and the market
        + # of house permits, Type of construction, City, postal code, province, New construction or renovation
    - Use cases examples
      * SEARCH BY FEATURE
        + What is the age of the account by customer for 2017?
        + What is the credit limit by customer class for 2017?
        + What is the avg payment terms for a residential contractor in 2017?
      * SEARCH BY PRICE
        + What is the avg selling price by type of houses by province in 2017?
        + What is the avg selling price for residential houses by province in 2017?
      * SEARCH BY MARKET TRENDS
        + What type of order a contractor orders?
        + How many residential houses were built by Postal code in 2017?
        + What are the sales by customers by SIC code in 2017 vs 2016?
        + What are the sales by product family in 2017?
        + What is the avg spend of electrical material according to a contractor?
        + How many permits for new construction of residential houses were awarded in 2017 by province?
        + How many permits were awarded in 2017 by province?
      * SEARCH BY TYPE OF MARKETS (Residential, commercial)
        + How many were awarded by cities, municipalities for 2017 by vertical market (resi, commercial)
        + What type of permits were awarded in 2017? i.e commercial, residential
        + What type of houses have been built in 2017 by province?
        + What is the avg sales by market segment?
        + What is the avg sales by type of customer class?
* SEARCH BY TYPE OF SIC CODE
  + What is the SIC code ratio for residential vs all SIC codes for 2017?
* SEARCH BY TYPE OF ROI
  + What is the P&L by market segment in 2017 for Quebec?
* SEARCH BY TYPE OF CUSTOMER
  + How many permits were awarded between contractors or consumers by postal code for 2017?
* Week 2: Data exploratory
  + Business Question
    - What could look like 2017 sales within product family, market segment & ship type?
  + Sub questions:
    - What could look like the projected sales for electrical products within residential market by product family in 2018?
    - What would be the best market segment to grow the sales in 2018?
    - What would be the best product family to grow our base line sales in 2018?
  + Data Exploratory
    - Four (4) datasets were exported
      * Nedco Sales 2016\_2017
      * SIC code from Nedco CRM
      * Housing starts from Stat Can
      * Housing type from Stat can
    - Challenges
      * Data complexity due to data cleaning
      * Lot of data to analyze from a time constraint perspective
  + Feedback from a technical perspective
    - Challenge to have the right coding
    - Challenge to resolve error within coding
    - Challenge to clean data before machine learning process
    - Challenge to learn and apply python coding
  + Feedback from a business perspective
    - The project was based on Nedco sales data only
  + 1st level of analysis: One variable
    - Define DF\_sales 2017 according to the # of transactions :
      * Including ‘NA’ : 144 468 transactions
      * Excluding ‘NA’ : 90 761 transactions



* + 2nd level of analysis using an histogram and box plot
    - Mean: avg sales order 1582
    - Std
    - Min per sales order:
    - Max





* + 3rd level of analysis: Sales by Product Family 2017 (Count)
    - CONNECT, FITT & OUT/BOXES 19730
    - LIGHTING 19532
    - WIRE & CABLE 16557
    - CONDUIT & ACCESSORIES 16240
    - DISTRIBUTION & FUSES 15281
    - DATACOM 14424
    - OTHER PRODUCTS 11668
    - WIRING DEVICES 10393
    - TOOLS 9692
    - HEATING PRODUCTS 5841
    - MOTOR CONTROLS/AUTOMATION 4364
    - SECURITY PRODUCTS 680
    - ALTERNATE ENERGY 30



* + 4th level of analysis: Multiple variables (numeric \* Numeric)
    - Validate the ship type and sales 2017 correlation’s

|  | **Ship Type** | **2017** |
| --- | --- | --- |
| **Ship Type** | 1.000000 | 0.173076 |
| **2017** | 0.173076 | 1.000000 |



* + 5th level of analysis: Multiple variables (Categorical x Numeric)
    - Box plot to show the distribution Market Segment Desc x sales 2017
    - Validate Ship type 1, 2,3 & product sales



* + 6th level of analysis: Multiple variables (Categorical x Categorical)
    - Validate sales between product family and market segment
    - Key questions to answer
      * Is the Product Family Score a good indicator of predictive sales?
      * Identify Product Family with high potential of sales?
      * As an sales executive, select a Product Family to predict sales?
* Week 3: Machine Learning
  + Feedback from a technical perspective
    - Challenging to have the right coding
    - Challenging to resolve error
    - Data was not as clean as it should be to process machine learning
  + Feedback from a business perspective
    - Challenge to find if market segment, Product family and ship type could be good indicators to establish a predictive sales model
  + Results
    - Due to time constraints and coding issues, the model was tested at level at Model Training / Evaluation - Using Split, but couldn’t figure out how to fix ValueError: could not convert string to float: '$1,628.00 '

|  |  |
| --- | --- |
| * Week 4: Cluster analysis   + Technical challenge   + Present the results |  |
|  |  |
|  |  |