Capstone NVR

PROBLEM:

1. Identify relationships, correlations, or predictive factors for:
   - The quality of the home prior to or post-settlement.
   - Budgeting and spending post-settlement.
2. ==Determine if specific products, geographic areas, vendors, or other factors impact quality or spending.==
3. Forecast the workload of warranty employees based on historical trends.
4. Analyze relationships or correlations between warranty employee ticket volume and customer experience.

OUR IDEA:

Vendor/Product Line Issue Detection Alert System
Goal: Identify when a specific product or vendor is causing a spike in ticket issues.
How:

- Use EDA findings on ProductLineDescription, VendorNumber, and ProblemCategory.
- Create a script to:

  - Aggregate ticket volume weekly/monthly by product line/vendor.
  - Compare to historical averages (e.g., 2 standard deviations above baseline).
  - Alert when issue volume spikes.

Benefit: Quicker vendor accountability and product recalls, lower long-term costs.

==Determine if specific products, geographic areas, vendors, or other factors impact quality or spending.==

Spending and quality – tie these two into our model

Performing well defined as amount spent and quality of product

Zoom meeting 4/25

- Focus on top 5 products
- Look at regions (problems in Florida and apply to the south)
    - Weather would play a role

Deliverables

- <mark>Create a model to determine if specific products, geographic areas, vendors, or other factors impact quality or spending.</mark>
    - *Spending and quality – tie these two into our model*
    - *'Performing well' is defined as amount spent and quality of product*
-

Want a dashboard where they can input their data (about spending and quality of home) and it shows them

Zoom meeting 4/29 with prof

- Recurring a lot? Find specific regions
- Look at different porducts in terms of tickets it creating or spending that occured
    - Anomily would find top regions to balance
- Simple sentences
- Input
- Processed
- Output
- What sell to client

- descriptive stats, find distribution, when things are not in balance

Team zoom meeting 5/1

- Target products and vendors and then compare

Isolation forest model

- Input
  - Percentage change ticket number
  - Percentage change invoice line amount
- Processed
  - Isolation forest model
- Output
  - Anomaly negative 1
    - Calculated patterns and unusual behavior (outliers)
  - Normal 1

Prophet model

- Input
  - Date time stamp
  - Y (ticket volume)
- Processed
  -
- Output

Zoom meeting with team 5/2

- Our model needs to output the top 5 products
- Create a model to rank them against each other to find the top 5

Input

- Invoice line amount
- phase number
- Assigned employee id
- Phase activity label
- Region

Processed

- Runs the data through decision tree
- Is the invoice amount greater than ...

- Then puts it in anomily or normal

Output

- Either anomaly (1)
- Or normal

Algorithm to find the top 5 products

- Aggregated and group them
- Analyse the anomaly and then rank them for top 5
  - Look at the amount spent and other features to rank the products

Find what the y is

What is an anomaly

- Could be
  - Found that when product is this product (4 tickets) its addidng $20000 on the expense side
  - Flagged it and found these are the problems

If we go with outliers

- Ticket number more than 20
  - Flag with y
  - 0 for tickets
- So system can take tickets and when each of them trigger
  - Then system alerts

For client

- Artifacts
- Group project paper
- Slides
- Code demo
- Dont go over 30mintues for presentation

Here is what we found in the data

Why we wanted to explore more

Original idea

Final product

Challenges we faced

Important features

Test and train split

20 bunch

Created a new model

1000

Trys to predict cost (y hat) -> attach to excel sheet

6 features/parameters

Count of

- Warranty issue count

- Warranty invoice count
- Vendor count
- Product name
- Ticket member
-

Linear regression

Random forest (comparison)

Logistic regression