**Welcome!**

Hello, and welcome to your home-assignment. This should take a couple of hours and not more than that. Please feel free to use Python, SQL, or R and your code could be in .txt, .R, or .py format. This test is mainly to access your technical ability to solve any data manipulation problems, understand the data through a business lens, and provide insights to a data scientist, business stakeholder, or a PM.

**Evaluation**

30% - Coding style: Ease of maintenance; terseness; use of best practices; leverage latest technologies / libraries / clever coding techniques; etc. Appropriate choice of 3rd party libraries or frameworks is encouraged.

20% - Data scrubbing: Steps taken to clean data and possibility of automating the cleansing step through scripts.

20% - Documentation: You may use any documentation style as long as it is clear what your code is doing.

30% - Analysis/ Data Engineering: Demonstrate a clear understanding of modeling, analysis, and data manipulation.

Feel free to ask any questions as you tackle challenge! Have Fun!

**Problem Statement:**

Create extra columns by extracting PFAMName to get configuration of the product as follows:

* 1. Column name as Processor-

Example: Device-1 i7/16/1TB GPU

Extract i7 in this row and populate in the column “Processor”.

There should be different types of processor like i5, i7, M etc.

* 1. Column name as RAM-

Example: Device-2 i7/16/1TB GPU

Extract 16 and populate as 16GB. There should be different types as 4GB, 8GB, 16GB etc.

* 1. Column name as HDD –

Example: Device-1 i7/16/1TB GPU

Extract 1TB and populate as 1TB. There should be different types as 1TB

* 1. Commercial Flag-

Example: Device-2 i5/8/256 Comm Demo

Extract Comm from Product hierarchy and Flag as 1 if found, 0 if not found.

* 1. Bundle Flag-

Example: Device-4 i5 4GB BlackTCBndl M1796

Extract a Bundle flag (Bndl) from Product Hierarchy and flag as 1 if found, 0 if not found.

1. Join the two provided csv files: one dimension and one fact table. Then, provide your understanding of the given data or any insights you gather. Feel free to use above flags you created above. This is where you can combine your data engineering and creativity together, which can help give data scientists a kickstart on the data they are working on.
2. Please provide insights through visualization and a summary slide of how you would present it to a business stakeholder. This will be where you would show your business acumen and data insights.

**Input files**

**There should be two csv files:**

1. **Fact\_table**

|  |  |
| --- | --- |
| dateid | Unique identifier for Date |
| CalendarDate | Calendar Date |
| CalendarQuarter | Quarter |
| YearMonth | Calendar Year and Calendar Month |
| SubsidiaryCode | Country ID |
| SubsidiaryName | Country Name |
| TPId | Partner ID on where items are sold |
| TPName | partner Name |
| ProductPartNbr | Item Number |
| SellThruQTY | Sell through Quantity - is what the Retailer sells to end customers |
| SellinQTY | Sell In Quantity -is what we sell to retailers |

1. **Dimensions**

|  |  |
| --- | --- |
| SubsidiaryCode | Country ID |
| SubsidiaryName | Country Name |
| TPId | Partner ID on where items are sold |
| TPName | partner Name |
| PFAMName | Part of Product Hierarchy Product Family Level |
| DeviceName | Part of Product Hierarchy Device Level (Highest grain) |
| ProductItemName | Product Name |
| ProductPartNbr | Item number |

Data Insights

1. **Gathering insights from the flags created**
   1. Majority of the business (sales) are coming through B2B (business-to-business) sales. Commercial sales are almost negligible.
   2. For the bundled products the sell thru qty/rate is higher (108.74) than sold which means, the demand for bundled products is high as compared to non-bundled one.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **commercial\_flag** | **bundle\_flag** | **sellinqty** | **sellthruqty** | **sell-thru-rate** |
| 0 | 0 | 512971 | 389484 | 75.93% |
| 0 | 1 | 132764 | 144368 | 108.74% |
| 1 | 1 | 0 | 6 | NA |
| 1 | 0 | 0 | 5672 | NA |

1. **Sales Trend**
   1. Overall sales went up in initial years from 2015 to 2016, there after it went down. Sell thru also followed the same trend, it spiked up from 2015 to 2016, the rate dropped a little going until 2018.
   2. This trend / correlation between sell and sell-thru shows that initially the products that were sold did not have enough demand in the market and the inventory went up in the market however the demand was not there. Eventually after mid 2017 the demand from customers crossed the sell quantity which was the break-even point where quantities sold equaled quantities sell – thru number.
   3. Overall product demand has been on a decline since 2018. Will dive deep into this direction.
2. **Which product hierarchy is selling the best or at decline**
   1. Below graph shows that initially when the sales were shooting up it was the Device-1 that was selling well until mid-2017.
   2. Beyond that Device-2 have done the best among all the product categories even as the sales have gone down significantly since mid-2018.
   3. Need to explore why device-1 or device-2 which were best sellers stopped selling? What were those factors?
3. **Which Device-1 sold the best**
   1. Drilling deeper into which device-1 had done the best. The config used is 128 GB, i5 and 4GB which is lower level config. Focus should be more on selling lower level config devices as they sell more 70% more than higher config devices.

|  |  |  |
| --- | --- | --- |
| Device Name | Qty Sold | Perc allocation |
| Device-1 128GB i5 4GB | 163663 | 49.01% |
| Device-1 256GB i5 8GB | 70389 | 21.08% |
| Device-1 128GB M 4GB | 22967 | 6.88% |
| Device-1 128GBM4GB DevNoPenSigCover Bndl | 19472 | 5.83% |
| Device-1 256GB i7 8GB | 17586 | 5.27% |
| Device-1 128GBi5 4GB DevNoPen Cover Bndl | 16077 | 4.81% |
| Device-1 256GB i7 16GB | 12332 | 3.69% |
| Device-1 128GB M 4GB No Pen Sustain | 7329 | 2.19% |
| Device-1 512GB i7 16GB | 2954 | 0.88% |
| Device-1 1TB i7 16GB | 512 | 0.15% |
| Device-1 128GB i5 4GB Demo | 337 | 0.10% |
| Device-1 256GB i5 8GB DevNoPen CoverBndl | 303 | 0.09% |

1. **Market trend**
   1. Shows that 128 GB and 256 GB HDD devices have dominated the market which means that demand was high for lower config devices
2. **By Country:**
   1. US sales numbers have dominated over Canada in the past

# Executive Summary-

**Insights-**

1. Sales have been on steep decline since mid-2018.
2. Until Mid-2017 when the sales were on rise, Device-1 had done the best among all other product categories. Between Mid-2017 and Mid-2018 Device-2 had done the best.
3. Looking at the type of devices that have sold the best in the past- the focus should be on bundled device packages and also on lower config devices (lower HDD, RAM and Processor).

**Next Steps-**

1. Need to do a market study of what device configurations came to the market in / after mid-2018 that influenced the sales negatively thereafter.
2. Explore the option of moving into B2C (business-to-customer) model as B2B (business-to-business) is not doing well and B2C sales are almost negligible.
3. What config – HDD/RAM/Processor are doing well in the market which are not being offered by our company.
4. In-depth study of the market competition and their products. Also consider what are the external factors that lead to decrease in sales.
5. External factors that could be influencing the sales for ex: businesses moving into cloud services so not ordering high config devices, economy in a slump etc.
6. Work with Trading Partners that have got us the highest sales in the past to revive the sales.