**Data Understanding**

First 5 rows

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | order\_id | order\_created\_at | order\_total\_price | order\_status | order\_discounts | product\_title | product\_sku | product\_quantity | product\_discount | product\_category | product\_deparments | product\_created\_at |
| 0 | 20570 | 2018-01-01T17:24:47.000Z | 449.97 | fulfilled | yes | Vertigo™ Semi-Hollow | M80-VHB-BLK | 1 | 0.00 | Electric Guitar Cases & Gig Bags | NaN | 2017-10-26T02:57:14.000Z |
| 1 | 20570 | 2018-01-01T17:24:47.000Z | 449.97 | fulfilled | yes | The Tick™ 2.0 Black | M80-TICK-V2-BLK | 1 | 0.00 | Pedals & Pedalboards Cases & Gig Bags | NaN | 2017-10-26T02:57:14.000Z |
| 2 | 20570 | 2018-01-01T17:24:47.000Z | 449.97 | fulfilled | yes | Pedalboard Lite (Silver) | PFX-PB-LT-SLV | 1 | 0.00 | Pedalboards | NaN | 2017-10-26T02:57:14.000Z |
| 3 | 20556 | 2018-01-02T10:56:19.000Z | 249.99 | fulfilled | yes | Vertigo™ Semi-Hollow | M80-VHB-BLK | 1 | 0.00 | Electric Guitar Cases & Gig Bags | NaN | 2017-10-26T02:57:14.000Z |
| 4 | 20554 | 2018-01-02T11:45:26.000Z | NaN | fulfilled | no | The FlyBy | EFX-FLY-BLK | 1 | 229.99 | DJ Equipment Cases & Gig Bags | NaN | 2017-10-26T02:57:14.000Z |

Last 5 rows

df.tail()

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | order\_id | order\_created\_at | order\_total\_price | order\_status | order\_discounts | product\_title | product\_sku | product\_quantity | product\_discount | product\_category | product\_deparments | product\_created\_at |
| 7555 | 379892 | 2021-03-07T09:22:59.000Z | 149.99 | fulfilled | yes | Sleeve Bass Guitar Case, Ash | M80-SEB-ASH | 1 | 0.00 | Bass Guitar Cases & Gig Bags | NaN | 2017-10-26T02:57:14.000Z |
| 7556 | 376309 | 2021-03-02T16:29:42.000Z | 35.00 | fulfilled | no | Vertigo Bass Guitar Case, Black | M80-VEB-BLK | 1 | 0.00 | Bass Guitar Cases & Gig Bags | NaN | 2017-10-26T02:57:14.000Z |
| 7557 | 376922 | 2021-03-03T05:08:00.000Z | 499.99 | fulfilled | yes | Pedalboard Carbon Large and Pro Accessory Case... | PFX-PBC-L-SLV-BDL | 1 | 0.00 | Pedalboards | NaN | 2020-01-14T18:48:13.000Z |
| 7558 | 377051 | 2021-03-03T16:47:59.000Z | 169.99 | fulfilled | yes | Tour Accessory Case 2.0, Black | M80-TOUR-V2-BLK | 1 | 0.00 | Pedals & Pedalboards Cases & Gig Bags | NaN | 2017-10-26T02:57:14.000Z |
| 7559 | 380829 | 2021-03-08T11:25:23.000Z | NaN | NaN | no | Classic FlyBy Backpack, Black | EFX-FLY-BLK | 1 | 229.99 | DJ Equipment Cases & Gig Bags | NaN | 2017-10-26T02:57:14.000Z |

Data start: 2018, Jan 1st

Data end: 2021, Mar 8th

Total 7560 rows of data

Check null values

<class 'pandas.core.frame.DataFrame'>

RangeIndex: 7560 entries, 0 to 7559

Data columns (total 12 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 order\_id 7560 non-null int64

1 order\_created\_at 7560 non-null object

2 order\_total\_price 3791 non-null float64

3 order\_status 7189 non-null object

4 order\_discounts 7560 non-null object

5 product\_title 7560 non-null object

6 product\_sku 7551 non-null object

7 product\_quantity 7560 non-null int64

8 product\_discount 7560 non-null float64

9 product\_category 7152 non-null object

10 product\_deparments 0 non-null float64

11 product\_created\_at 7152 non-null object

dtypes: float64(3), int64(2), object(7)

memory usage: 708.9+ KB

**Action after data understanding phase**

After data understanding phase, required action for drop, transformation, or use.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Col | Column Name | Non Nulls | Data Type | Remarks | Action | Use |
| 0 | order\_id | 7560 | int64 | Unique ID | Retain | Use |
| 1 | order\_created\_at | 7560 | object | date | Reformat into year, month, day of week | Use |
| 2 | order\_total\_price | 3791 | float64 | Not every order has price |  | Keep for later use |
| 3 | order\_status | 7189 | object | 1. fulfilled 7102 (out of 7560) 2. partial 87 | Impute those null values with “non\_fulfilled” | Use |
| 4 | order\_discounts | 7560 | object | 1. no 4542 2. yes 3018 |  | Keep for later use |
| 5 | product\_title | 7560 | object | Classic FlyBy Backpack, Black 270  Classic Tick Accessory Case 2.0, Black 225  Vertigo™ Electric Guitar 174  Classic Dual Electric Guitar Case, Black 165  Vertigo Electric Guitar Case, Black 159  Name: product\_title, Length: 268, dtype: int64 | Got 268 types of products, therefore only use the top 10 or 20 products for analysis | Use as identifier per product |
| 6 | product\_sku | 7551 | object | EFX-FLY-BLK 459  M80-TICK-V2-BLK 373  M80-VEG-BLK 356  M80-EG-BLK 277  M80-2G-BLK 276  ...  B-M80-AC-BLK 1  ZP1-L-BLK 1  B-M80-VEG-GRY 1  50-K61-FOAM 1  TEX-BLK 1  Name: product\_sku, Length: 118, dtype: int64 | Got 118 types of product\_sku, therefore only use the top 10 product\_sku for analysis | TBD, Initially drop |
| 7 | product\_quantity | 7560 | int64 | 1 5675  2 612  4 477  5 214  3 187  8 123  10 109  6 46  12 34  20 21  16 20  15 13  7 6  9 5  25 4  11 3  13 2  28 2  24 2  14 2  30 1  42 1  19 1  Name: product\_quantity, dtype: int64 | Most items ordered are only single item only (5675).  Items ordered twice occurred 612 times. | Use |
| 8 | product\_discount | 7560 | float64 | 0.00 6722  229.99 243  329.99 115  249.99 109  49.99 55  0.02 50  189.99 48  149.99 45  89.99 45  169.99 21  199.99 16  179.99 9  209.99 7  79.99 7  Mostly no discounts | Most items are sold without discounts (6722 out of 7560). | TBD,  Initially drop  Keep for later use |
| 9 | product\_category | 7152 | object | Electric Guitar Cases & Gig Bags 1804  Bass Guitar Cases & Gig Bags 1034  Pedals & Pedalboards Cases & Gig Bags 1017  Acoustic Guitar Cases & Gig Bags 762  Guitar Straps 602  DJ Equipment Cases & Gig Bags 585  Drum & Percussion Cases & Gig Bags 544  Pedalboards 381  Wallets 168  Keyboard Cases & Gig Bags 91  Gifts 90  Ukulele Cases & Gig Bags 57  Speaker Stands 17  Name: product\_category, dtype: int64 | Maybe analysis by Category is more fruitful, TBD | Use |
| 10 | product\_deparments | 0 | float64 | No data | Null | Drop |
| 11 | product\_created\_at | 7152 | object | 2017-10-26T02:57:14.000Z 6484  2018-06-28T00:43:55.000Z 140  2019-05-12T22:30:50.000Z 139  2018-02-01T02:03:39.000Z 124  2018-05-31T18:01:09.000Z 93  2019-06-25T18:45:53.000Z 86  2020-07-22T20:16:48.000Z 25  2018-12-31T07:15:33.000Z 22  2020-01-14T18:48:13.000Z 17  2020-01-09T19:16:17.000Z 17  2020-05-18T22:16:51.000Z 5  Name: product\_created\_at, dtype: int64 | Date the product is created in database | Drop |

**Business objectives for Project**

1. Minimize unsold inventory, ideally, only a minimum stock is kept to cater to order from customers. Large unsold inventory is undesirable because it reduces cash flow and working capital. Unsold stock also experiences loss in value due to depreciations.
2. Ideally, all orders are fulfilled with zero stockout, because unfulfilled orders represent loss of sales for the company. Unfulfilled order also incurs ill will from customers.

Suitable metrics shall be proposed for measuring unfulfilled orders and unsold inventory. Based on past three years data, unfulfilled order is 5%.

To assist with the business objectives, accurate prediction of the demand from customers will help maintaining an “ideal” inventory for the products.

**Approach**

1. We will make use of machine learning to build a “best” model to predict the demand from customers based on the last 3 years of product order data.
2. The predictions from this model will be used to assist in decision making for
   1. Replenish depleted stock,
   2. Interventions to stimulate sales like offering discounts

These predictions shall be used to update an interactive dashboard for the top 10 (or 20) products. When stock fall below a specific level, a flag shall be raised to indicate action required to replenish stock. On the other hand, if stock remains stubbornly high for a period of time, trigger may be raised for sales stimulus actions like offering sales discount.

Diagram

Description automatically generated

This dashboard shall be simple and yet helpful to the management in visualizing the performance of the product inventory.

**Modelling**

An array of modelling techniques will be used:

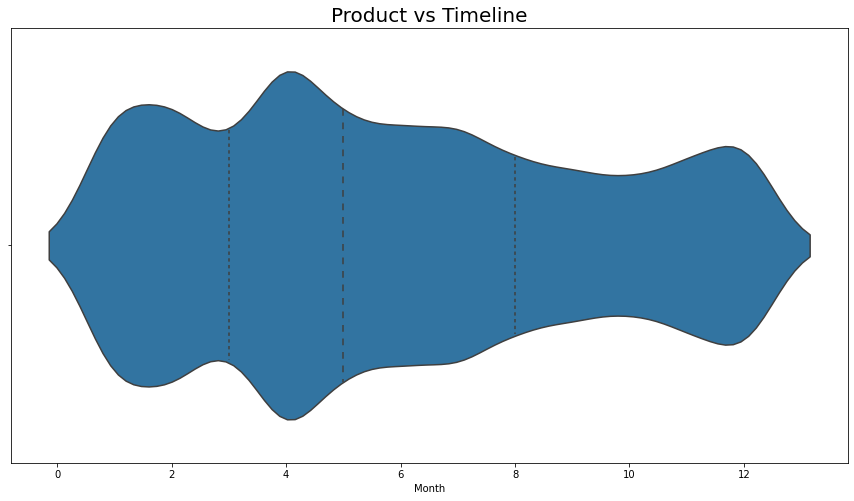
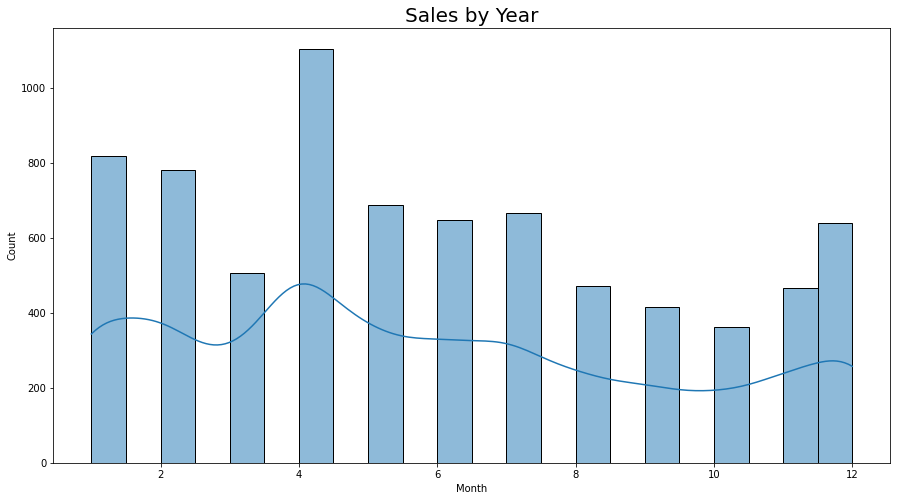
1. Cluster analysis shall be used to find out if some products are associated with other products in the sales order.
2. Associative analysis shall be performed to find out the co-relationships of some products
3. Time series analysis will be performed to analyse the sales of products against time. This will help to identify any peak and lull in product sales.
4. Regression modelling, single or multiple variables, to find out the ideal curve to fit the current product demand.
5. May include overlay of special events to explain peaks or lulls, eg.: school holidays, musical festivals and sales promotions.

**Inference**

After the best model is created, a hybrid reasoning system shall be designed to incorporate the results from the modelling to generate intervention actions. The predictions from the models will be input into a scoring (inference) engine. The scoring engine will interpret the prediction into intervention actions. The inference may be hard rules or may incorporate a fuzzy logic system inside.

A picture containing text, whiteboard

Description automatically generated

Visualizations from data

