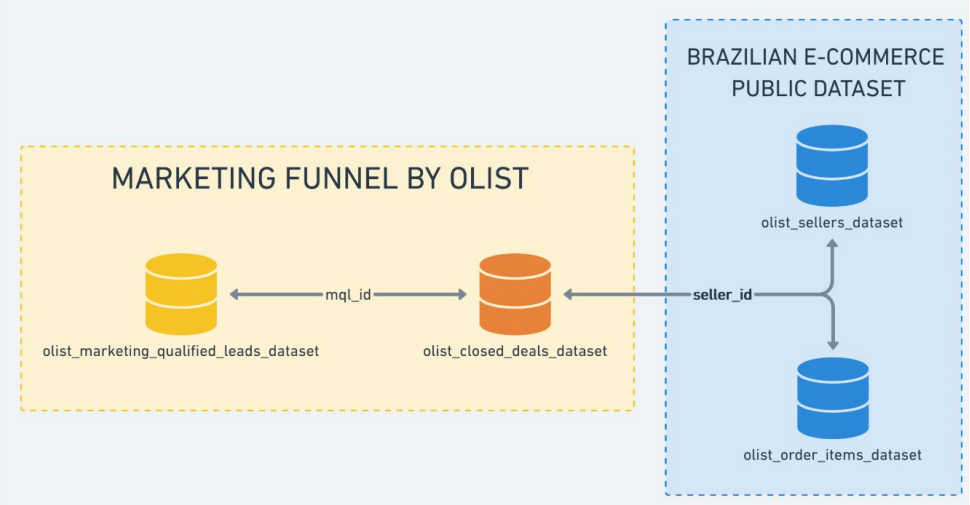
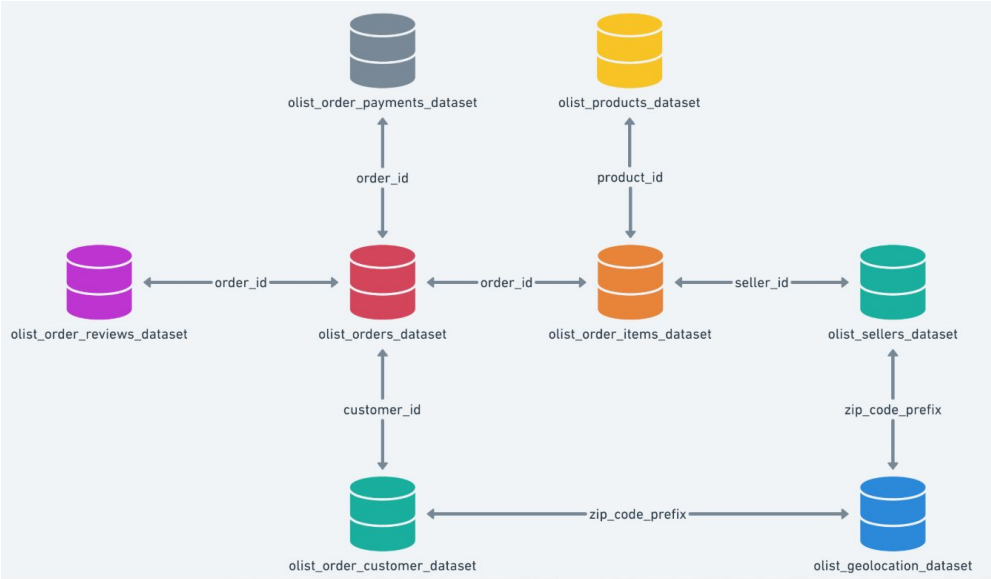
Marketing Funnel By Olist

DATA

Marketing funnel dataset from sellers who filled-in requests of contact to sell their products on Olist Store. It consists of 2 major datasets - **Marketing Qualified Leads dataset** - the dataset generated after the leads (either reseller or manufacturer) fills in data at the landing page to sell their products on Olist store. **Closing Deals Dataset** - the lead is contacted by a Sales Development Representative and gathers information about the product,business segment, type etc.

Merging with Brazilian datasets - Sellers data, Orders and Products for our analysis.




olist_order_payments_dataset
olist_products_dataset
order_id
product_id
order_id
order_id
seller_id
olist_order_reviews_dataset
olist_orders_dataset
olist_order_items_dataset
olist_sellers_dataset
customer_id
zip_code_prefix
zip_code_prefix
olist_order_customer_dataset
olist_geolocation_dataset

XSV

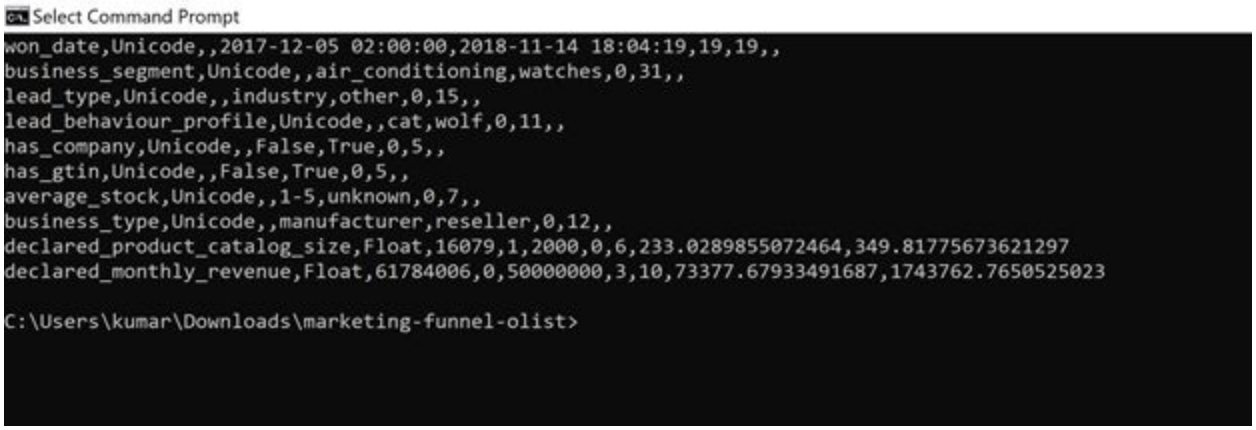
XSV- A command line program offering high performance accessibility to the Datasets where you can quickly jump into the statistics of the data like its frequency,count, length, its structure and do quick operations like join,slicing,indexing on any simplified dataset.

Headers -

```
C:\Users\kumar\Downloads\marketing-funnel-olist>xsv headers olist_closed_deals_dataset.csv
1   mql_id
2   seller_id
3   sdr_id
4   sr_id
5   won_date
6   business_segment
7   lead_type
8   lead_behaviour_profile
9   has_company
10  has_gtin
11  average_stock
12  business_type
13  declared_product_catalog_size
14  declared_monthly_revenue

C:\Users\kumar\Downloads\marketing-funnel-olist>
```

Status



Frequency -

```
average_stock,(NULL),776
average_stock,5-20,22
average_stock,50-200,15
average_stock,1-5,10
average_stock,20-50,8
average_stock,200+,7
average_stock,unknown,4
business_type,reseller,587
business_type,manufacturer,242
business_type,(NULL),10
business_type,other,3
declared_product_catalog_size,(NULL),773
declared_product_catalog_size,100.0,9
declared_product_catalog_size,50.0,7
declared_product_catalog_size,300.0,5
declared_product_catalog_size,400.0,4
declared_product_catalog_size,20.0,4
declared_product_catalog_size,10.0,3
declared_product_catalog_size,1000.0,3
declared_product_catalog_size,30.0,2
declared_product_catalog_size,40.0,2
declared_monthly_revenue,0.0,797
declared_monthly_revenue,100000.0,5
declared_monthly_revenue,20000.0,3
declared_monthly_revenue,25000.0,3
declared_monthly_revenue,30000.0,3
declared_monthly_revenue,10000.0,3
declared_monthly_revenue,5000.0,2
declared_monthly_revenue,60000.0,2
declared_monthly_revenue,250000.0,2
declared_monthly_revenue,300000.0,2

C:\Users\kumar\Downloads\marketing-funnel-olist>xsv frequency olist_closed_deals_dataset.csv
```

Slicing -

```
C:\Users\kumar\Downloads\marketing-funnel-olist>xsv slice olist_closed_deals_dataset.csv -s 48 | xsv table
mql_id                           seller_id                        sdr_id                           sr_id                            won_date            busines
s_segment               lead_type       lead_behaviour_profile  has_company  has_gtin  average_stock  business_type  declared_product_catalog_size  declared_mont
hly_revenue
c1e598764329cc9c377ef1d029be8ceb  d4dcd3a31f5f19aea2f258a568517a4f  fdb16d3cbbeb5798f2f66c4096be026d  060c0a26f19f4d66b42e0d8796688490  2018-02-22 13:18:26  food_dr
ink                     industry                                                                    manufacturer                                    0.0
9cca8d684240e24dd459f2d439fae30c  28760737b4b82d4479ddff8db5e9f71f  9e4d1098a3b0f5da39b0bc48f9876645  4ef15afb4b2723d8f3d81e51ec7afefe  2018-02-27 20:04:12  pet
                        online_medium    cat                                                        manufacturer                                    0.0
0d99b130d767e50e22b528261ffe5550  0dccb6973ef48a00fbcccfe907181ab0  34d40cdaf94010a1d05b0d6212f9e909  a8387c01a09e99ce014107505b92388c  2018-06-11 12:04:11  food_su
pplement                online_beginner  cat                                                        reseller                                        0.0
7c9ac4388867d4cc3f1cf9c05ad7e944  751e274377499a8503fd6243ad9c56f6  9749123c950bf8363ace42cb1c2d0815  4ef15afb4b2723d8f3d81e51ec7afefe  2018-07-30 14:34:51  audio_v
ideo_electronics        online_big       cat                                                        reseller                                        0.0
2c5d0035798bffadf8f67ef0af84d183  93a55b2252867fd7df54c78cbd5c6d95  56bf83c4bb35763a51c2baab501b4c67  fbf4aef3f6915dc0c3c97d6812522f6a  2018-04-02 14:20:33  constru
ction_tools_house_garden  offline         cat                                                        reseller                                        0.0
e5860ae00103b3869a25d940345bf0fd  1185ae3e5ac0463c854dad54fc8fc9e1  34d40cdaf94010a1d05b0d6212f9e909  495d4e95a8cf8bbf8b432b612a2aa328  2018-04-13 12:50:35  bags_ba
ckpacks                 industry         cat                                                        manufacturer                                    0.0
014c073ef65e4a7ef73acd61cf606164  2646baaf662d4d92ac48f047e35db92d  9d12ef1a7eca3ec58c545c678af7869c  4ef15afb4b2723d8f3d81e51ec7afefe  2018-03-23 19:06:54  constru
ction_tools_house_garden  online_medium    eagle                                                      reseller                                        0.0
3a91a097116d6c8b32d41e0c81dd1e9c  e433f5fd4050e3352b5d83522b7fe24b  a8387c01a09e99ce014107505b92388c  9e4d1098a3b0f5da39b0bc48f9876645  2018-04-26 03:00:00  constru
ction_tools_house_garden  online_medium    cat                                                        manufacturer                                    0.0
```

TRIFACTA

In this assignment Trifacta has been used for three critical tasks which are essential for any data wrangling process. These tasks are Cleaning the data(i.e. Handling null values), Formatting the data in the desired format(eg. Yy:mm:dd to mm:dd:yyyy) and merging of required files to get a unified cleaned dataset.

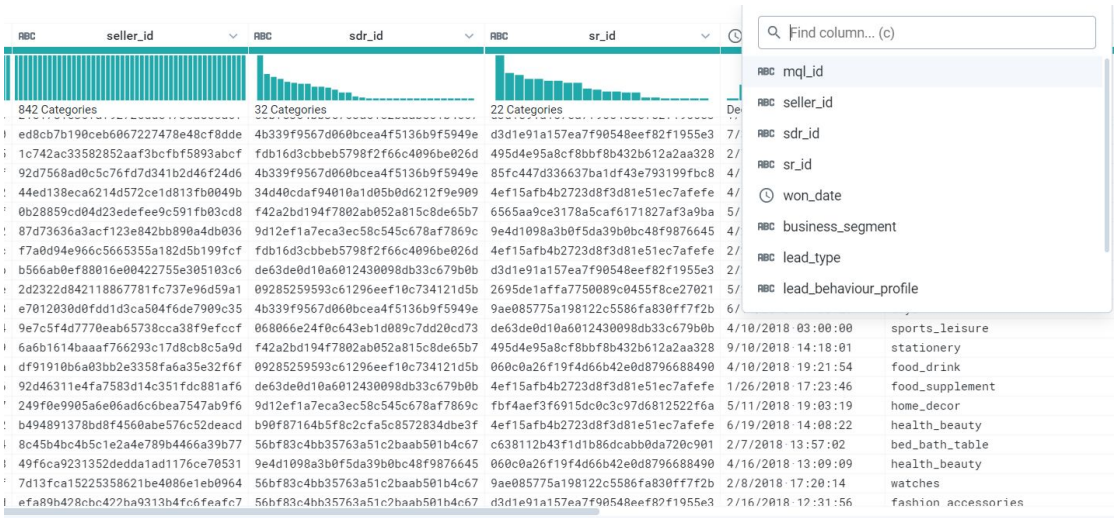
The major advantage of using Trifacta over other data wrangling tools is its ease of use. With a few drag-drops and mouse clicks we can be done with the above mentioned essential tasks in a matter of minutes. But, in order to perform these tasks, we have to create our own recipe within a dataflow wherein all these operations take place.

In the image below, two csv files have been imported, namely ‘olist_closed_deals_dataset.csv’ and ‘olist_marketing_qualified_leads.csv’. Two recipes have been created for each of the files respectively for cleaning and formatting purposes. Once these operations are performed we are adding another step in the ‘olist_closet_deals_dataset’ recipe to merge the two files.

The image below shows the various operations that we are performing within the recipe:

```
1  Change date format of won_date to M/d/yyyy HH:mm:ss
2  Set has_company to IFMISSING($col, NULL())
3  Delete has_company
4  Delete has_gtin
5  Delete average_stock
6  Delete declared_product_catalog_size
7  Set lead_behaviour_profile to IFMISSING($col, 'N\/A')
8  Set business_type to IFMISSING($col, 'N\/A')
9  Set lead_type to IFMISSING($col, 'N\/A')
10 Inner join with olist_marketing_qualified_leads_dataset on mql_id == mql_id
11 Set business_segment to IFMISSING($col, 'N\/A')
12 Delete mql_id1
```

Finally, after all the operations within the recipe have been executed we will get our desired dataset which will look something similar to the image below.



This forms the basis of our source data wherein we get to know the number of closed deals which will play an important role in analysing our data.

PANDAS

Considering the Brazilian E-commerce datasets - loading the following files:

```
In [4]: #olist order items
        order_items = pd.read_csv('olist_order_items_dataset.csv')

In [5]: #olist sellers dataset
        sellers = pd.read_csv('olist_sellers_dataset.csv')

In [6]: #olist products dataset
        products = pd.read_csv('olist_products_dataset.csv')

In [7]: #product category name translation
        prod_name = pd.read_csv('product_category_name_translation.csv')

In [8]: #olist orders dataset
        orders = pd.read_csv('olist_orders_dataset.csv')

In [9]: #olist orders reviews dataset
        reviews = pd.read_csv('olist_order_reviews_dataset.csv')
```

The orders datasets includes - Orders, Order Items and Order Reviews -

Order_Items -

```
In [93]: order_items.columns

Out[93]: Index(['order_id', 'order_item_id', 'product_id', 'seller_id',
                'shipping_limit_date', 'price', 'freight_value'],
               dtype='object')

In [47]: order_items.shape

Out[47]: (112650, 7)
```

Null value check -

```
In [48]: order_items.isnull().sum()

Out[48]: order_id              0
         order_item_id         0
         product_id            0
         seller_id             0
         shipping_limit_date   0
         price                 0
         freight_value         0
         dtype: int64
```

No null values

Orders dataset - It consists of 99441 rows and 8 columns

```
In [95]: orders.columns

Out[95]: Index(['order_id', 'customer_id', 'order_status', 'order_purchase_timestamp',
                'order_approved_at', 'order_delivered_carrier_date',
                'order_delivered_customer_date', 'order_estimated_delivery_date'],
               dtype='object')
```

```
In [52]: orders.shape

Out[52]: (99441, 8)
```

Null check -

```
In [53]: orders.isnull().sum()

Out[53]: order_id                            0
         customer_id                         0
         order_status                        0
         order_purchase_timestamp            0
         order_approved_at                 160
         order_delivered_carrier_date     1783
         order_delivered_customer_date    2965
         order_estimated_delivery_date       0
         dtype: int64
```

Dropping the rows with null values -

```
In [54]: orders.dropna(subset=['order_approved_at','order_delivered_carrier_date','order_delivered_customer_date',
                               'order_estimated_delivery_date'],inplace=True)
         orders.isnull().sum()

Out[54]: order_id                        0
         customer_id                     0
         order_status                    0
         order_purchase_timestamp        0
         order_approved_at               0
         order_delivered_carrier_date    0
         order_delivered_customer_date   0
         order_estimated_delivery_date   0
         dtype: int64
```

Date formatting - formatting the date time stamp to just date.

```
orders['order_approved_at'] = pd.to_datetime(orders['order_approved_at']).dt.date
orders['order_delivered_carrier_date'] = pd.to_datetime(orders['order_delivered_carrier_date']).dt.date
orders['order_delivered_customer_date'] = pd.to_datetime(orders['order_delivered_customer_date']).dt.date
orders['order_estimated_delivery_date'] = pd.to_datetime(orders['order_estimated_delivery_date']).dt.date
```

Order Reviews - The order reviews consists of 1,00,000 and 7 columns

```
In [194]: reviews.columns

Out[194]: Index(['review_id', 'order_id', 'review_score', 'review_creation_date',
                 'review_answer_timestamp'],
                dtype='object')
```

```
In [161]: reviews.shape

Out[161]: (100000, 7)
```

Null check -

```
In [162]: reviews.isnull().sum()

Out[162]: review_id                     0
          order_id                      0
          review_score                  0
          review_comment_title      88285
          review_comment_message    58247
          review_creation_date          0
          review_answer_timestamp       0
          dtype: int64
```

Dropping columns - review_comment_title and review_comment_message (column/data in Portuguese)

```
reviews.drop(["review_comment_title","review_comment_message"],axis=1,inplace=True)
```

```
reviews.isnull().any()

]: review_id                 False
   order_id                  False
   review_score              False
   review_creation_date      False
   review_answer_timestamp   False
   dtype: bool
```

Merging the Orders dataset with order_items,

```
In [166]: ord = orders.merge(order_items, on='order_id', how='left')
          ord.head(10)
```

```
In [196]: ord.columns

Out[196]: Index(['order_id', 'customer_id', 'order_status', 'order_purchase_timestamp',
                 'order_approved_at', 'order_delivered_carrier_date',
                 'order_delivered_customer_date', 'order_estimated_delivery_date',
                 'order_item_id', 'product_id', 'seller_id', 'shipping_limit_date',
                 'price', 'freight_value'],
                dtype='object')

In [197]: ord.shape

Out[197]: (110180, 14)
```

Then merging the ord with reviews -

```
In [168]: order = ord.merge(reviews, on='order_id', how='left')
          order.head(10)
```

```
In [171]: order.shape

Out[171]: (95831, 18)
```

Writing to Orders csv file --

```
In [172]: order.to_csv('Order.csv',index = False)
```

Finally we get the 95,831 rows and 18 columns from Orders, which is imported into Salesforce analytics.

Seller -

```
In [140]: sellers.head(5)
```

Out[140]:

	seller_id	seller_zip_code_prefix	seller_city	seller_state
0	3442f8959a84dea7ee197c632cb2df15	13023	campinas	SP
1	d1b65fc7debc3361ea86b5f14c68d2e2	13844	mogi guacu	SP
2	ce3ad9de960102d0677a81f5d0bb7b2d	20031	rio de janeiro	RJ
3	c0f3eea2e14555b6faeea3dd58c1b1c3	4195	sao paulo	SP
4	51a04a8a6bdcb23deccc82b0b80742cf	12914	braganca paulista	SP

```
In [141]: sellers.shape
```

Out[141]: (3095, 4)

```
In [142]: sellers.isnull().any()
```

Out[142]:
```
seller_id                False
seller_zip_code_prefix   False
seller_city              False
seller_state             False
dtype: bool
```

```
In [143]: sellers.to_csv('Sellers_data_EDA.csv',index = False)
```

Cleaned data and writing it to csv file.

Product -

Performed data cleaning and Merged two different product datasets - Product and Product Category names (basically it had names in Portuguese and english names).

```
In [182]: product = prod_name.merge(products, on='product_category_name', how='inner')
          product.head(10)
```

Out[182]:

	product_category_name	product_category_name_english	product_id
0	beleza_saude	health_beauty	e3e020af31d4d89d2602272b315c3f6e
1	beleza_saude	health_beauty	c5d8079278e912d7e3b6beb48ecb56e8
2	beleza_saude	health_beauty	36555a2f528d7b2a255c504191445d39
3	beleza_saude	health_beauty	e586ebb6022265ae1eea38f46ffe3ead
4	beleza_saude	health_beauty	75b4372e69a42f8ae1d908c076f547b2
5	beleza_saude	health_beauty	3569d4374a919941a50f57371b1dc93d
6	beleza_saude	health_beauty	3a6a0247ced9dcb444b46caafdcdd368
7	beleza_saude	health_beauty	adf591c625cb265c12bc6749d3a2f757
8	beleza_saude	health_beauty	50556c630443502c11acde1c320fe278
9	beleza_saude	health_beauty	88d2c501ec765f5d7e8038fa6aab0e62

Major products (names in english)--

```
In [184]: product['product_category_name_english'].unique()

Out[184]: array(['health_beauty', 'computers_accessories', 'auto', 'bed_bath_table',
       'furniture_decor', 'sports_leisure', 'perfumery', 'housewares',
       'telephony', 'watches_gifts', 'food_drink', 'baby', 'stationery',
       'tablets_printing_image', 'toys', 'fixed_telephony',
       'garden_tools', 'fashion_bags_accessories', 'small_appliances',
       'consoles_games', 'audio', 'fashion_shoes', 'cool_stuff',
       'luggage_accessories', 'air_conditioning',
       'construction_tools_construction',
       'kitchen_dining_laundry_garden_furniture',
       'costruction_tools_garden', 'fashion_male_clothing', 'pet_shop',
       'office_furniture', 'market_place', 'electronics',
       'home_appliances', 'party_supplies', 'home_confort',
       'costruction_tools_tools', 'agro_industry_and_commerce',
       'furniture_mattress_and_upholstery', 'books_technical',
       'home_construction', 'musical_instruments',
       'furniture_living_room', 'construction_tools_lights',
       'industry_commerce_and_business', 'food', 'art',
       'furniture_bedroom', 'books_general_interest',
       'construction_tools_safety', 'fashion_underwear_beach',
       'fashion_sport', 'signaling_and_security', 'computers',
       'christmas_supplies', 'fashio_female_clothing',
       'home_appliances_2', 'books_imported', 'drinks', 'cine_photo',
       'la_cuisine', 'music', 'home_comfort_2',
       'small_appliances_home_oven_and_coffee', 'cds_dvds_musicals',
       'dvds_blu_ray', 'flowers', 'arts_and_craftmanship',
       'diapers_and_hygiene', 'fashion_childrens_clothes',
       'security_and_services'], dtype=object)
```

Writing to Product csv file--

```
In [185]: product.to_csv('Product.csv',index = False)
```

Importing these 3 files - Orders, Products and Sellers into Einstein Analytics.

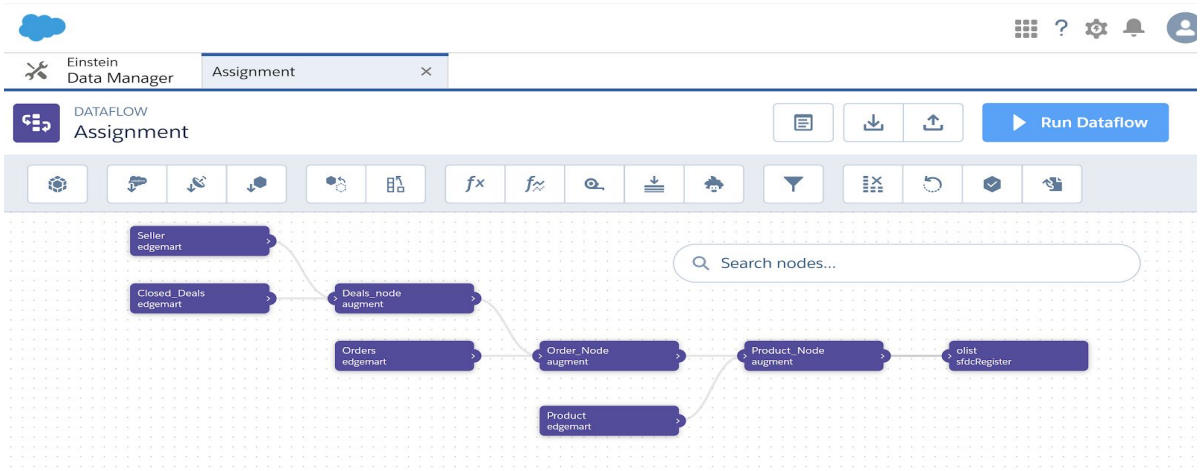
EINSTEIN ANALYTICS - Dashboard

https://na174.salesforce.com/analytics/wave/dashboard?assetId=0FK6g000000BpgxGAC&orgId=00D6g000006NVYL&loginHost=na174.salesforce.com&urlType=sharing&pageId=f1e04ab4-d573-4b86-aa85-5c8c14f8bf6e&savedViewId=8wk6g000000g1BeAAI&analyticsContext=analyticsTab

Dimensions - Product, Business Segment, Business Type, Seller State and City, Lead Behavior profile, Origin

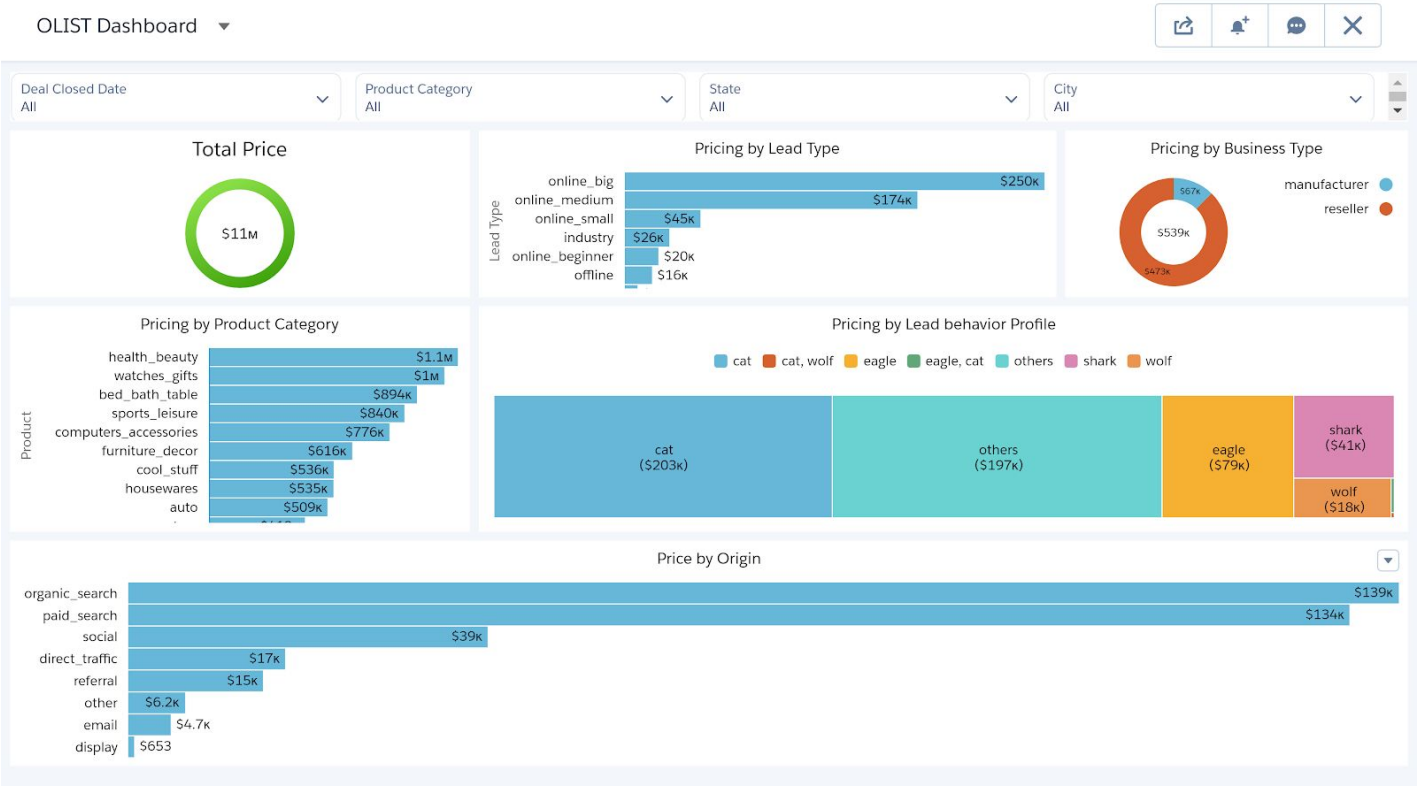
Measures - Pricing, freight value and review score

Data Flow -

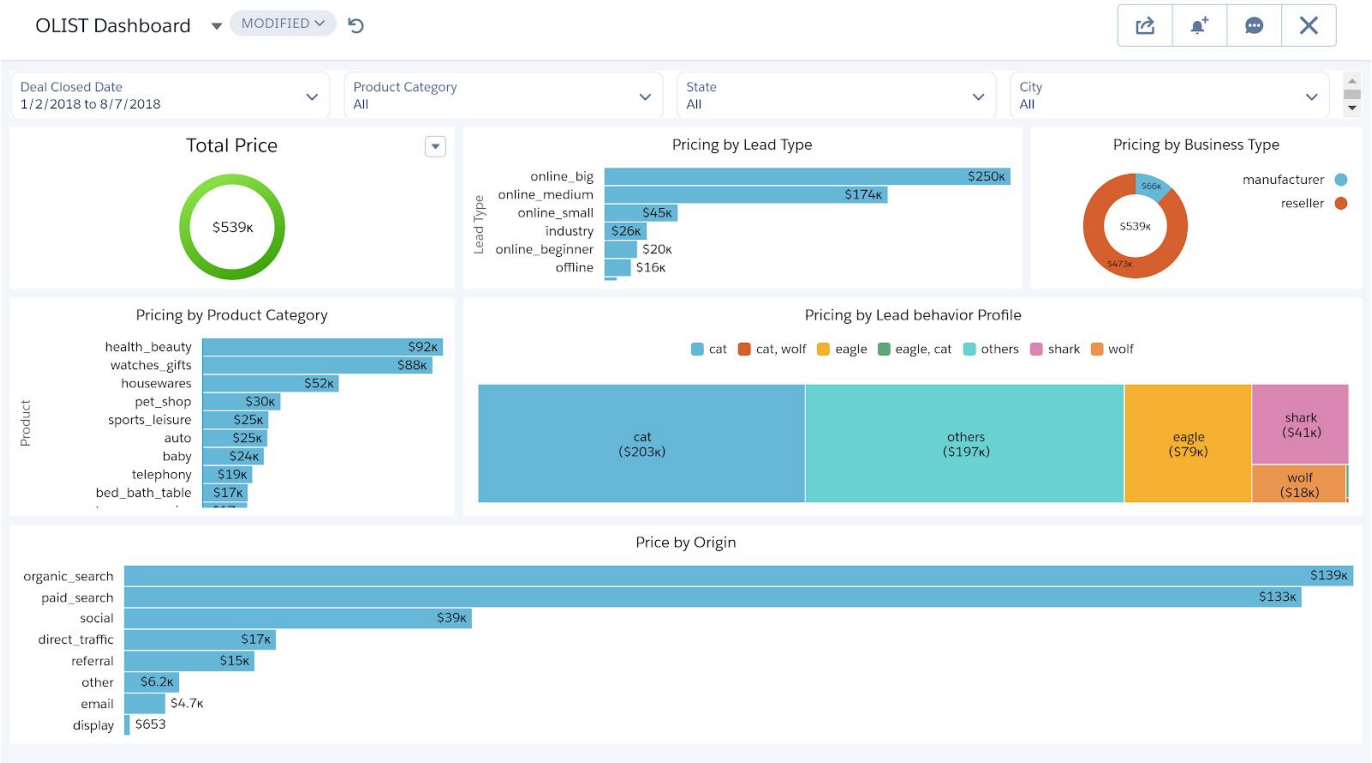


Pricing Dashboard - with date, state, city and Product filters .

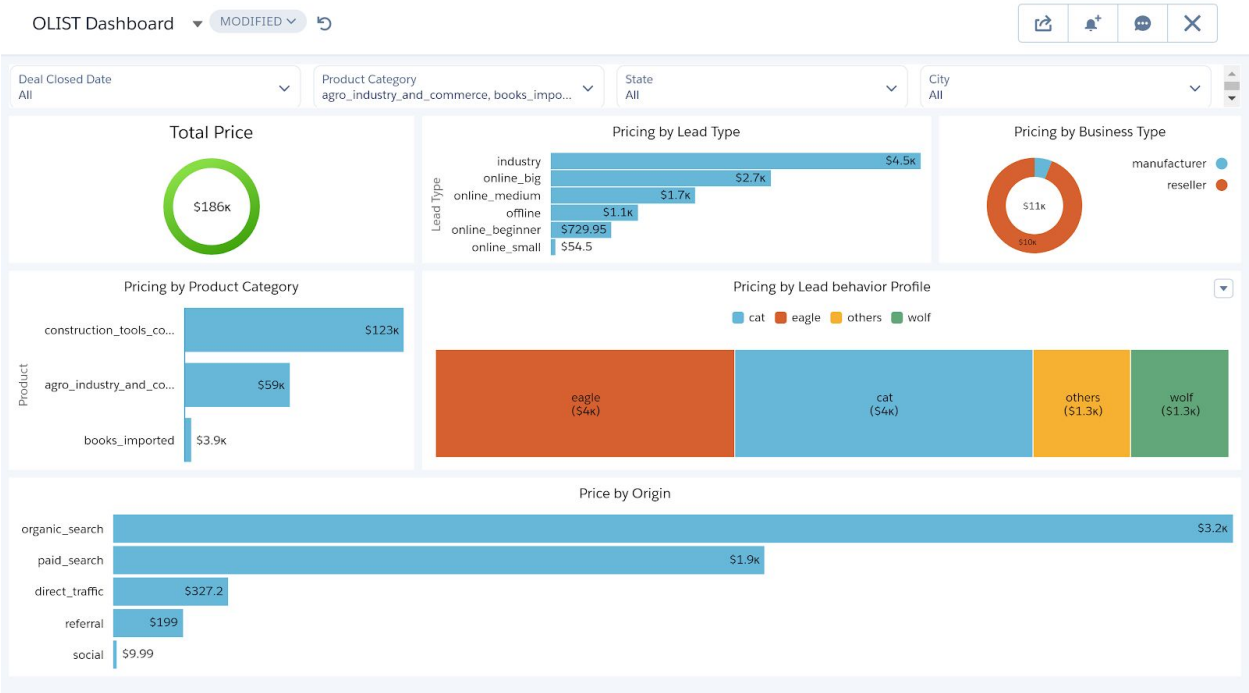
It basically consists of pricing by Product category, Lead Type, Business Type, Lead Behavior profile, by origin and Total price.



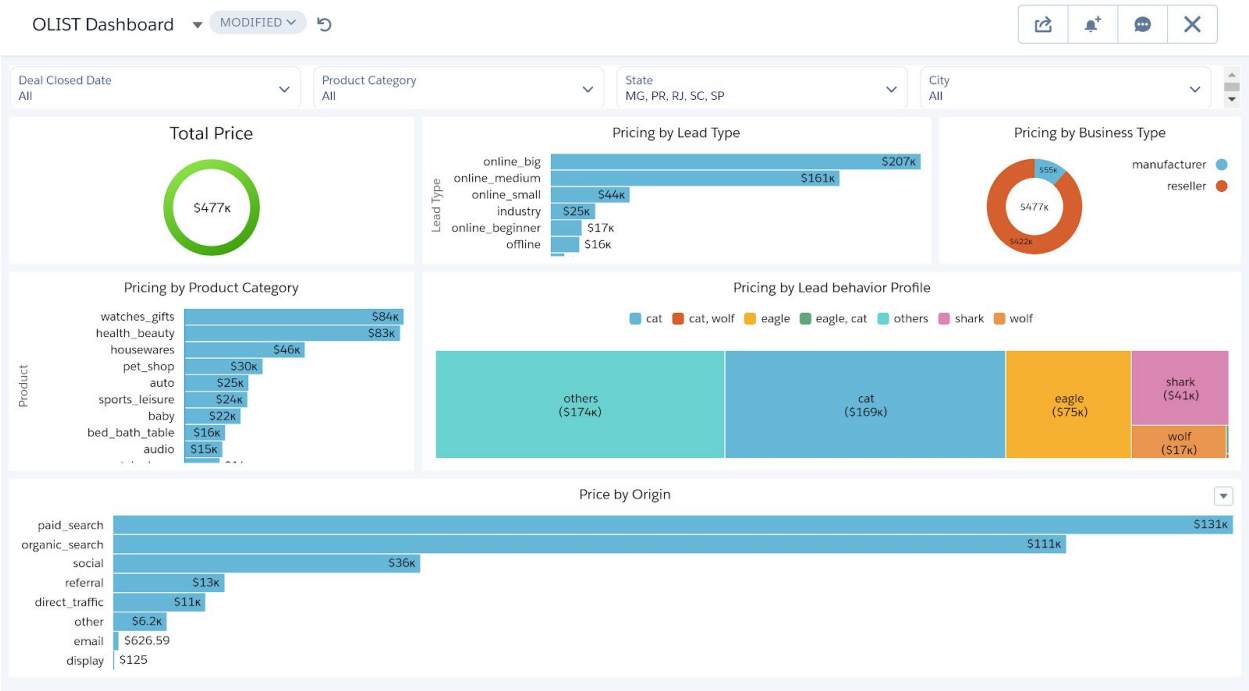
Based on Date filter - from 1/2/2018 to 8/7/2018



Based on Product filter -



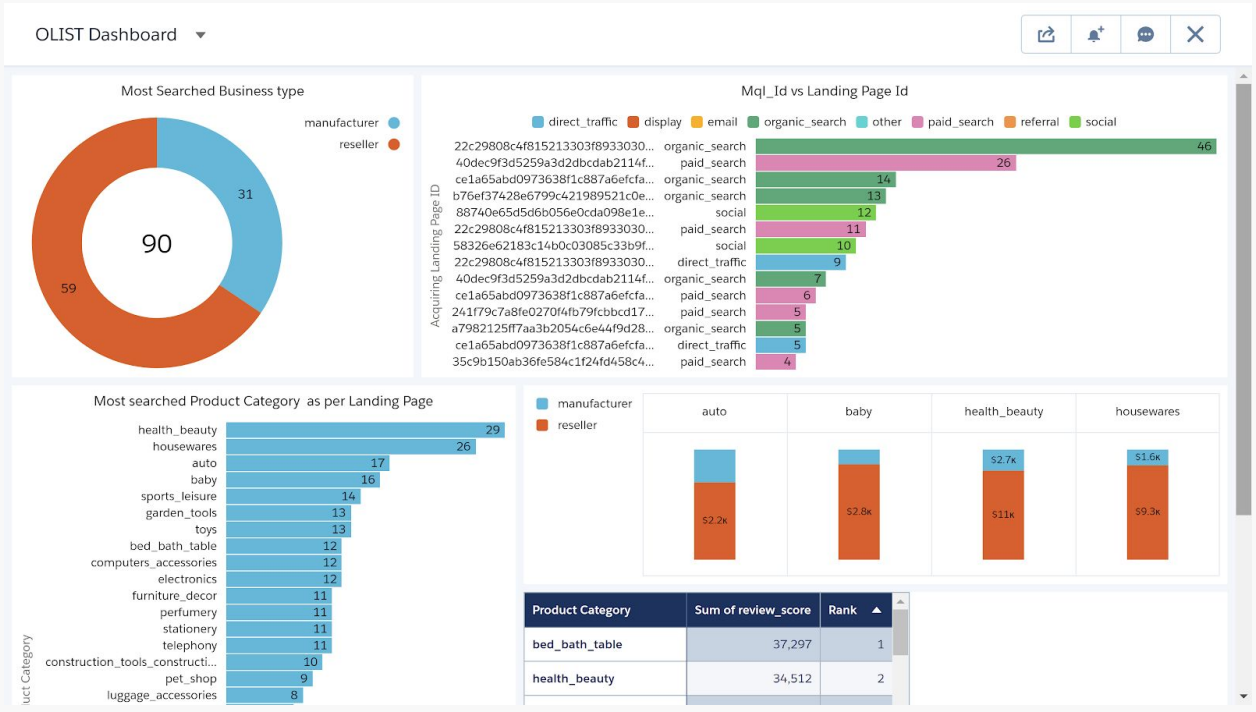
Based on state filter- MG, PR, RJ, SC, SP.



SEARCH

Search Criteria On the basis of maximum Search via Landing Page Id -

Most Searched product, Business type and Lead type.



Continued - - Product review score(sum of scores) and the rank based on customer provided review score for the orders.


OLIST Dashboard
a7982125ff7aa3b2054c6e44f9d28... organic_search 5
ce1a65abd0973638f1c887a6efcfa... direct_traffic 5
35c9b150ab36fe584c1f24fd458c4... paid_search 4
Most searched Product Category as per Landing Page
Product Category
health_beauty 29
housewares 26
auto 17
baby 16
sports_leisure 14
garden_tools 13
toys 13
bed_bath_table 12
computers_accessories 12
electronics 12
furniture_decor 11
perfumery 11
stationery 11
telephony 11
construction_tools_constructi... 10
pet_shop 9
luggage_accessories 8
cool_stuff 7
food 7
consoles_games 6
home_appliances 6
small_appliances 6
construction_tools_lights 5
home_appliances_2 5
home_construction 5
watches_gifts 5
air_conditioning 4
art 4
construction_tools_safety 4
fixed_telephony 4
furniture_living_room 4
manufacturer
reseller
auto
baby
health_beauty
housewares
$2.2K
$2.8K
$2.7K
$11K
$1.6K
$9.3K
Product Category
Sum of review_score
Rank
bed_bath_table 37,297 1
health_beauty 34,512 2
sports_leisure 30,612 3
furniture_decor 27,735 4
computers_accessories 26,592 5
housewares 24,124 6
watches_gifts 20,494 7
telephony 15,145 8
garden_tools 15,122 9