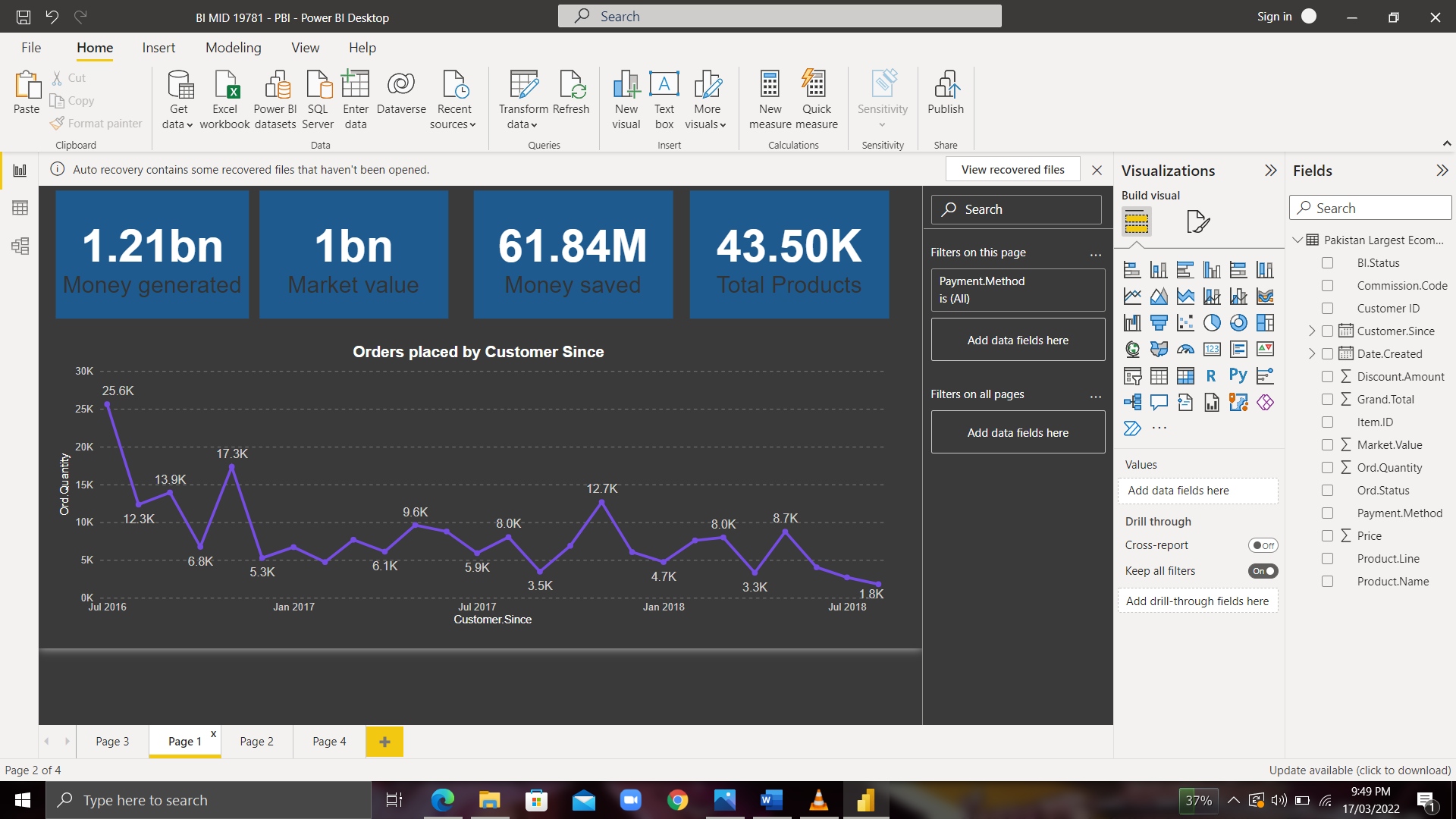
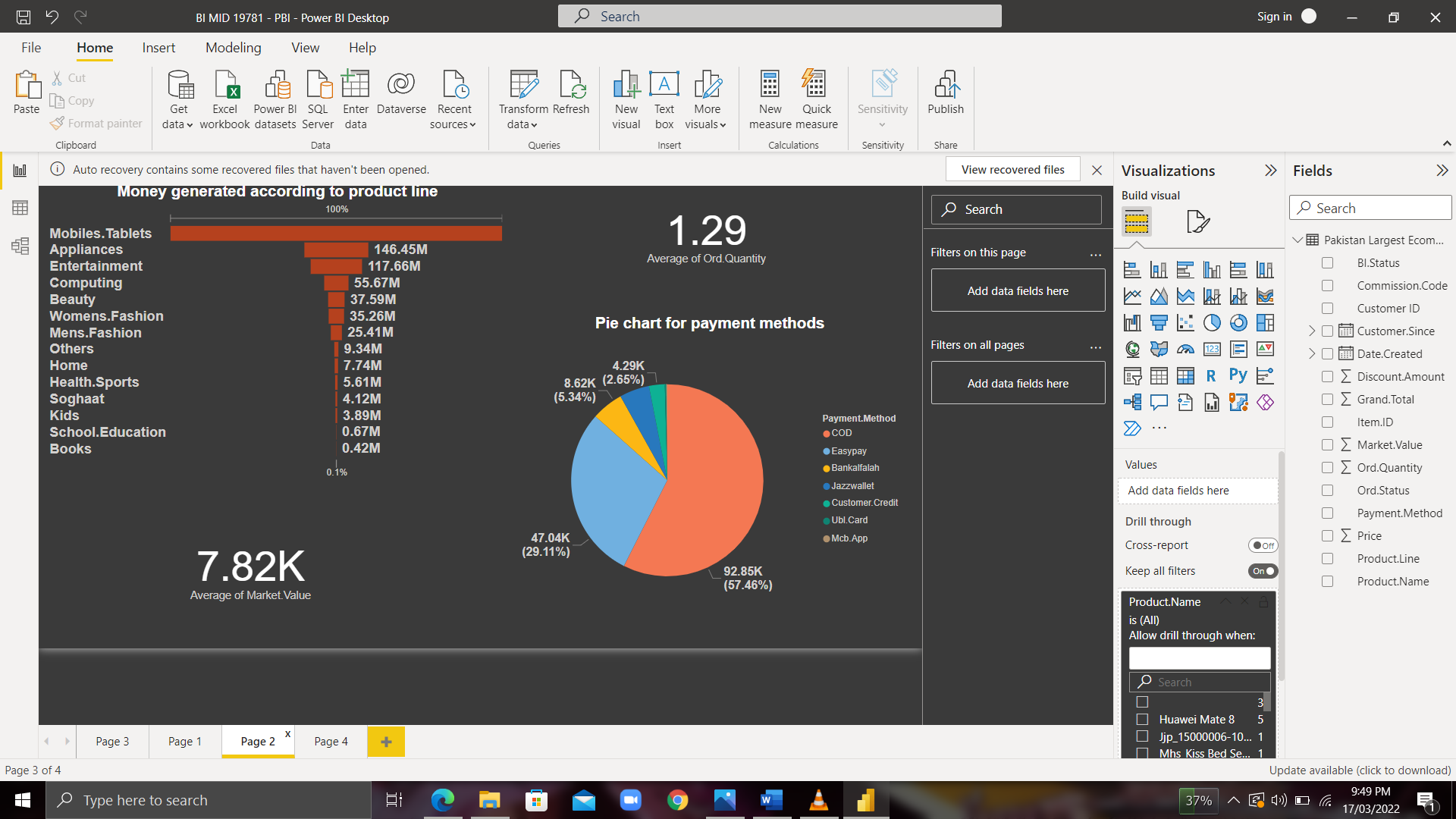
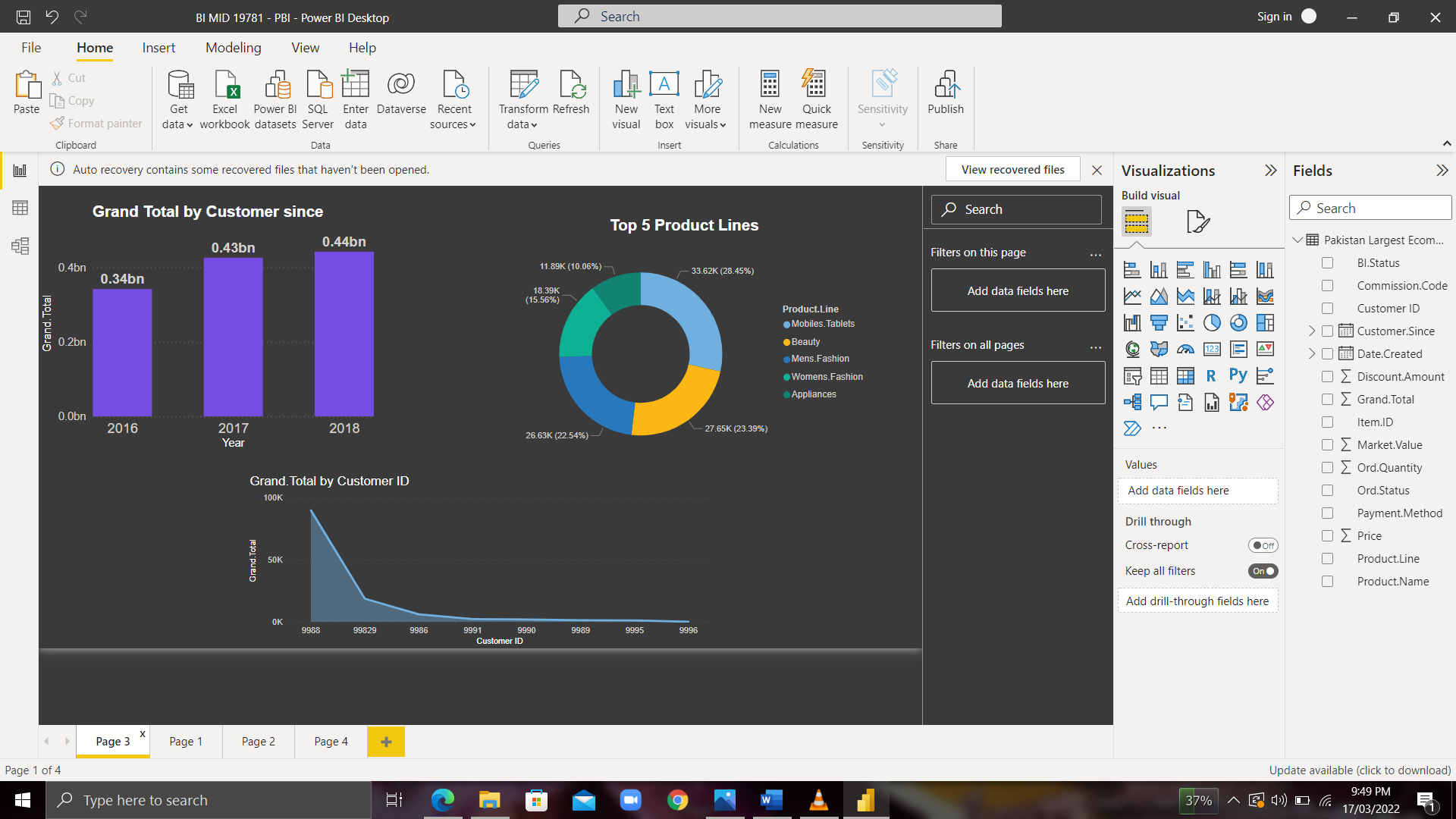
BI Assignment -Midterm

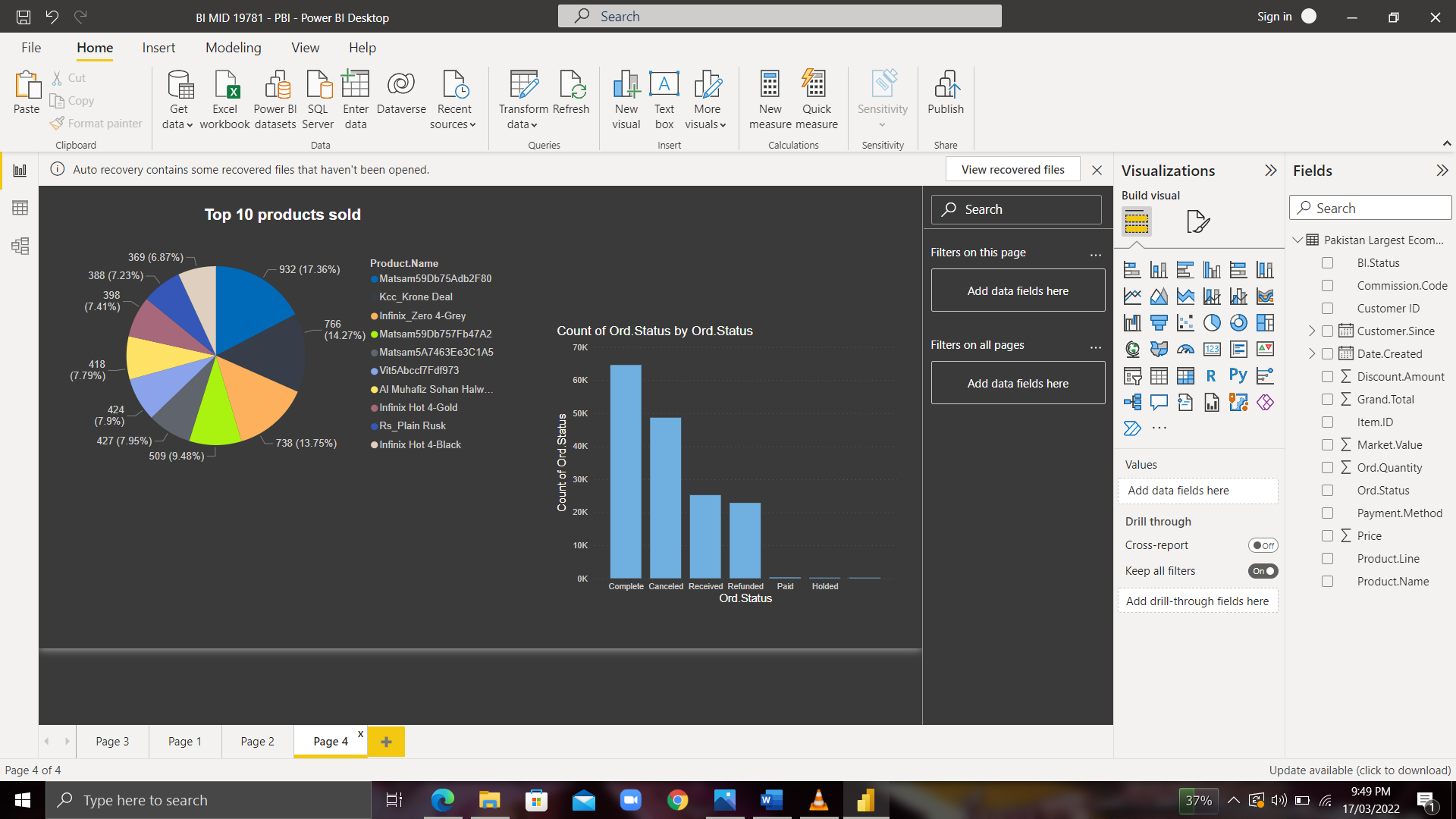
Roshaan Tariq khan-19781

**Data set used:** largest ecommerce dataset

* **Problem statement:** Do older customers order more products than newer ones? And should the ecommerce provide more attention and perks to them?
* **Steps in cleaning:**
* Renamed columns (ALL)
* Changed data types of columns to appropriate ones
* Replaced names of values in columns to appropriate ones
* Rounded off decimal values to two dp
* Removed unnecessary columns such as Fy, M-Y Working date, as these columns have the same values as date created which Ive kept.
* Removed increment id as it does not seem useful to our analysis.
* Added custom column of Grand.Total, which computes the total properly keeping in check the discount.
* Combined some values into one such as COD & cashatdoorstep to COD. Similarly combined similar values such as all Easypay variants into just Easypay and same with jazzcash.
* Renamed /N in commission. Code to NOcomm so its clearer.
* Removed blank rows and rows with errors.
* Removed 400k null values from Price column as filling them with mean or median would make the price inaccurate and since the column was our main column.
* Cleaned dirty values from columns
* Removed negative values in Grand.Total column by using formulae to make neg values to zero.
* Kept multiple payment methods as they all seem valid.
* Financesettlement replaced with COD as greatest mode.
* Mygateway replaced with customer. credit
* Payaxis combined with easypay as it has greatest mode for online payments.
* Apg combined with customer. credit as it pertains to some strategy related to number of the patient’s visit.
* Combined some order status options which were more similar such as order pending, holded to holded.
* Renamed product line’s names to shorter clearer ones.
* Removed rows with price value of 0 as it cannot be analyzed.
* Used capitalize first word
* Dimensions(time):
* Date variable: created at, Customer since
* Dimensions(normal):
* Product.Name
* Product.Line
* Payment.Method
* Order.Status
* Bi status
* Commission code
* KPI’s:
* Grand.Total – As this shows the total money generated / profit earned
* Order.Status- This shows how many orders have succeded and how many have failed indicating performance of the ecommerce platform.
* Customer since- older customers coming back to the system means the system is doing well to serve its consumers.
* Disclaimer:
* Some columns were kept to see if they come in handy later on. If they were not included in the charts , means they were not as necessary or , did not take priority over others.
* Questions:
* Analyze KPI’s across Product.Name
* Analyze KPI’s across Product.Line
* Analyze KPI’s across Payment.Method
* Analyze KPI’s across Order.Status
* Time dimension:

-What are the temporal trends of the KPI’s across Customer.Since.

* Chart Analysis:
* My problem statement asked the question if older customers tend to buy more products. As seen in the cumulative curve, yes older customers seem to trust the ecommerce system more and hence buy more products and hence spend more.
* The funnel chart shows that mobile and tablets are the leading category for generating the most money(grand total).(Seems obvious as the most expensive products were amongst this category)
* The pie chart shows Top payment methods used .
* Since Grand total is our key column here , one of the above graphs showcases grand total generated by customer since whereas the second shows top customers who have spent the most in the ecommerce platfrom and are the consumers the strategists should focus on keeping and providing perks.
* The third chart shows top 5 product lines dominating the ecommerce market.



* The first pie chart shows top 10 products sold. The ecommerce platform should focus on advertising to these brands more so they can offer more growth to the plaform.
* The bar graph uses another KPI (order status) to show the count for each order status. Since the complete status is the highest , means the system is right on track. The count for cancelled status brings worry , as they need to focus on using more frequent payment methods used to make the order successful.