

Exploratory Data Analysis with R

Microsoft Power BI - Part III

Xuemao Zhang
East Stroudsburg University

November 30, 2022

Visualizing northwind data

- NorthWind Dashboard in Power BI https://github.com/asyaparfenova/dashboard_NorthWind_Database/blob/main/PowerBI/PowerBI_dashboard.md

Transform Data

- Import the data from the northwind database
 - ▶ Choose the 3 tables only and click Transform Data

Navigator

Display Options ▾

localhost:5432: northwind [14]

- ☒ public.categories
- ☐ public.customer_customer_demo
- ☐ public.customer_demographics
- ☐ public.customers
- ☐ public.employee_territories
- ☐ public.employees
- ☒ public.order_details
- ☐ public.orders
- ☐ public.products
- ☐ public.region
- ☐ public.shippers
- ☒ public.suppliers
- ☐ public.territories
- ☐ public.us_states

public.suppliers

supplier_id	company_name	contact_name
1	Exotic Liquids	Charlotte Cooper
2	New Orleans Cajun Delights	Shelley Burke
3	Grandma Kelly's Homestead	Regina Murphy
4	Tokyo Traders	Yoshi Nagase
5	Cooperativa de Quesos 'Las Cabras'	Antonio del Valle Saavedra
6	Mayumi's	Mayumi Ohno
7	Pavlova, Ltd.	Ian Devling
8	Specialty Biscuits, Ltd.	Peter Wilson
9	PB Knäckebröd AB	Lars Peterson
10	Refrescos Americanas LTDA	Carlos Diaz
11	Heli Süßwaren GmbH & Co. KG	Petra Winkler
12	Plutzer Lebensmittelgroßmärkte AG	Martin Bein
13	Nord-Ost-Fisch Handelsgesellschaft mbH	Sven Petersen
14	Formaggi Fortini s.r.l.	Elio Rossi
15	Norske Meierier	Beate Vileid
16	Bigfoot Breweries	Cheryl Saylor
17	Svensk Sjöföda AB	Michael Björn
18	Aux joyeux ecclésiastiques	Guyène Nodier
19	New England Seafood Cannery	Robb Merchant
20	Leka Trading	Chandra Leka
21	Lyngbysild	Niels Petersen
22	Zaanse Snoepfabriek	Dirk Luchte
23	Karkki Oy	Anne Heikkinen

Select Related Tables

Load Transform Data Cancel

Transform Data - Table order_details

- Expand table orders

- ▶ We can always come back to update the data transform by clicking Home -> Transform data

The screenshot shows the Tableau interface with the 'public.orders' table expanded. The table has columns: 1.2 unit_price, 1.2 quantity, and 1.2 discount. The 'order_date' column is highlighted with a green circle and a red arrow pointing to it with the text 'You can select this as well'. The 'Use original column name as prefix' checkbox is also circled in red.

	1.2 unit_price	1.2 quantity	1.2 discount
1	11	14	
2	42	9.80000019	
3	72	34.79999924	
4	14	18.60000038	
5	51	42.40000153	
6	41	7.699999809	
7	51	42.40000153	
8	65	16.79999924	
9	22	16.79999924	
10	57	15.60000038	
11	65	16.79999924	
12	20	64.80000305	
13	33	2	
14	60	27.20000076	
15	31	10	
16	39	14.39999962	
17	49	16	
18	24	3.599999905	
19	55	19.20000076	
20	74	8	
21	2	15.19999981	
22	16	13.89999962	

Search Columns to Expand

- ☒ order_id
- ☐ customer_id
- ☐ employee_id
- ☒ order_date
- ☐ required_date
- ☒ shipped_date
- ☐ ship_via
- ☐ freight
- ☐ ship_name
- ☐ ship_address
- ☐ ship_city
- ☐ ship_region
- ☐ ship_postal_code
- ☐ ship_country
- ☒ public.customers
- ☒ public.employees
- ☐ public.order_details
- ☒ public.shippers

☒ Use original column name as prefix

OK Cancel

Transform Data - Table order_details

- Then we expand table customers

Formula bar: `= Table.ExpandRecordColumn(public_order_details, "public.orders", {"ord`

	order_date	shipped_date	public.customers
1	0	7/4,	
2	0	7/4,	
3	0	7/4,	
4	0	7/5,	
5	0	7/5,	
6	0	7/8,	
7	0.150000006	7/8,	
8	0.150000006	7/8,	
9	0.050000001	7/8,	
10	0.050000001	7/8,	
11	0	7/8,	
12	0.050000001	7/9,	
13	0.050000001	7/9,	
14	0	7/9,	
15	0	7/10,	
16	0	7/10,	
17	0	7/10,	
18	0.150000006	7/11,	
19	0.150000006	7/11,	
20	0	7/11/1996	7/23/1996 Value

Search Columns to Expand

- ☐ (Select All Columns)
- ☐ customer_id
- ☒ company_name
- ☒ contact_name
- ☒ contact_title
- ☐ address
- ☒ city
- ☐ region
- ☐ postal_code
- ☒ country
- ☐ phone
- ☐ fax
- ☐ public.customer_customer_demo
- ☐ public.orders

☐ Use original column name as prefix

OK Cancel

Transform Data - Table order_details

- Then we expand table employees

fx = Table.ExpandRecordColumn("#Expanded public.orders", "public.customers", {"company_name", "con

	contact_title	city	country	public.employees
1	Accounting Manager	Reims		
2	Accounting Manager	Reims		
3	Accounting Manager	Reims		
4	Marketing Manager	Münster		
5	Marketing Manager	Münster		
6	Accounting Manager	Rio de Janeiro		
7	Accounting Manager	Rio de Janeiro		
8	Accounting Manager	Rio de Janeiro		
9	Sales Agent	Lyon		
10	Sales Agent	Lyon		
11	Sales Agent	Lyon		
12	Accounting Manager	Charleroi		
13	Accounting Manager	Charleroi		
14	Accounting Manager	Charleroi		
15	Accounting Manager	Rio de Janeiro		
16	Accounting Manager	Rio de Janeiro		
17	Accounting Manager	Rio de Janeiro		
18	Owner	Bern		
19	Owner	Bern		
20	Owner	Bern		
21	Sales Manager	Genève		
22	Sales Manager	Genève		
23	Sales Manager	Genève		

Search Columns to Expand

- (Select All Columns)
- ☒ employee_id
- ☒ last_name
- ☒ first_name
- ☒ title
- ☐ title_of_courtesy
- ☐ birth_date
- ☐ hire_date
- ☐ address
- ☐ city
- ☐ region
- ☐ postal_code
- ☐ country
- ☐ home_phone
- ☐ extension
- ☐ photo
- ☐ notes
- ☐ reports_to

☐ Use original column name as prefix

OK Cancel

Switzerland Value

Transform Data - Table order_details

- Then we expand table shippers and rename the column as shipper_company_name

The screenshot shows the RStudio interface with a data table and an open 'Expand Columns' dialog box. The table has columns: id, last_name, first_name, title, and public.shippers. The dialog box is titled 'Expand Columns' and contains a search bar 'Search Columns to Expand'. Below the search bar, there is a list of columns to expand: (Select All Columns), shipper_id, company_name (selected), phone, and public.orders. At the bottom of the dialog, there is a checkbox 'Use original column name as prefix' which is unchecked. The 'OK' button is highlighted in yellow.

	id	last_name	first_name	title	public.shippers
1	5	Buchanan	Steven		
2	5	Buchanan	Steven		
3	5	Buchanan	Steven		
4	6	Suyama	Michael		
5	6	Suyama	Michael		
6	4	Peacock	Margaret		
7	4	Peacock	Margaret		
8	4	Peacock	Margaret		
9	3	Leverling	Janet		
10	3	Leverling	Janet		
11	3	Leverling	Janet	Sales Representative	Value
12	4	Peacock	Margaret	Sales Representative	Value
13	4	Peacock	Margaret	Sales Representative	Value
14	4	Peacock	Margaret	Sales Representative	Value
15	3	Leverling	Janet	Sales Representative	Value

Transform Data - Table order_details

= Table.RenameColumns("#Expanded public.shippers",{{"company_name.1", "shipper_company_name"}})					
ee_id	last_name	first_name	title	shipper_company_name	
1	5 Buchanan	Steven	Sales Manager	Federal Shipping	
2	5 Buchanan	Steven	Sales Manager	Federal Shipping	
3	5 Buchanan	Steven	Sales Manager	Federal Shipping	
4	6 Suyama	Michael	Sales Representative	Speedy Express	
5	6 Suyama	Michael	Sales Representative	Speedy Express	
6	4 Peacock	Margaret	Sales Representative	United Package	
7	4 Peacock	Margaret	Sales Representative	United Package	

Transform Data - Table order_details

- Then we remove table products

Query Settings

pers",{{"company_name.1", "shipper_company_name"}})

title	shipper_company_name	public.products
Sales Manager	Federal Shipping	Value
Sales Manager	Federal Shipping	Value
Sales Manager	Federal Shipping	Value
Sales Representative	Speedy Express	Value
Sales Representative	Speedy Express	Value
Sales Representative	United Package	Value
Sales Representative	United Package	Value
Sales Representative	United Package	Value
Sales Representative	Speedy Express	Value
Sales Representative	Speedy Express	Value
Sales Representative	Speedy Express	Value
Sales Representative	United Package	Value
Sales Representative	United Package	Value
Sales Representative	United Package	Value
Sales Representative	United Package	Value
Sales Representative	United Package	Value
Sales Representative	United Package	Value
Sales Representative	United Package	Value
Sales Manager	United Package	Value
Sales Manager	United Package	Value
Sales Manager	United Package	Value
Sales Representative	Federal Shipping	Value
Sales Representative	Federal Shipping	Value

Context menu options:

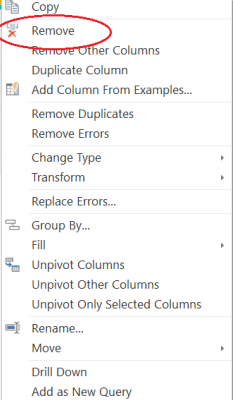
- Copy
- Remove**
- Remove Other Columns
- Duplicate Column
- Add Column From Examples...
- Remove Errors
- Replace Errors...
- Fill
- Unpivot Columns
- Unpivot Other Columns
- Unpivot Only Selected Columns
- Rename...
- Move
- Drill Down
- Add as New Query

Transform Data - Table categories

- Remove column picture

Source{[Schema="public",Item="categories"]}[Data]

	category_name	description	picture
1	Beverages	Soft drinks, coffees, teas, beers, and ales	Binary
2	Condiments	Sweet and savory sauces, relishes, spread...	Binary
3	Confections	Desserts, candies, and sweet breads	Binary
4	Dairy Products	Cheeses	Binary
5	Grains/Cereals	Breads, crackers, pasta, and cereal	Binary
6	Meat/Poultry	Prepared meats	Binary
7	Produce	Dried fruit and bean curd	Binary
8	Seafood	Seaweed and fish	Binary



The screenshot shows a context menu for the 'picture' column. The menu items are: Copy, Remove (circled in red), Remove Other Columns, Duplicate Column, Add Column From Examples..., Remove Duplicates, Remove Errors, Change Type, Transform, Replace Errors..., Group By..., Fill, Unpivot Columns, Unpivot Other Columns, Unpivot Only Selected Columns, Rename..., Move, Drill Down, and Add as New Query.

Transform Data - Table categories

- Expand table products

= Table.RemoveColumns(public_categories,{"picture"})

	1.2 category_id	A ^B category_name	A ^B description
1		1 Beverages	Soft
2		2 Condiments	Swee
3		3 Confections	Dess
4		4 Dairy Products	Chee
5		5 Grains/Cereals	Brea
6		6 Meat/Poultry	Prep
7		7 Produce	Dried
8		8 Seafood	Seaw

Search Columns to Expand

☒ Expand ☐ Aggregate

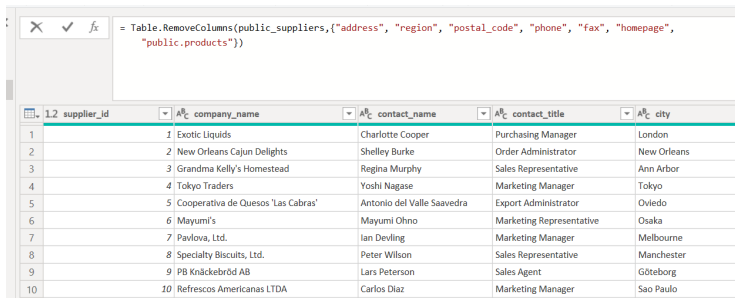
- ☐ (Select All Columns)
- ☒ product_id
- ☒ product_name
- ☒ supplier_id
- ☐ category_id
- ☐ quantity_per_unit
- ☐ unit_price
- ☐ units_in_stock
- ☐ units_on_order
- ☐ reorder_level
- ☐ discontinued
- ☐ public.categories
- ☐ public.order_details
- ☐ public.suppliers

☐ Use original column name as prefix

Transform Data - Table suppliers

- Keep only the following columns

- ▶ company_name
- ▶ contact_name
- ▶ contact_title
- ▶ city
- ▶ country



The screenshot shows a data transformation interface. At the top, a formula bar contains the expression: `= Table.RemoveColumns(public_suppliers,{"address", "region", "postal_code", "phone", "fax", "homepage", "public.products"})`. Below the formula bar is a table with 10 rows and 6 columns. The columns are labeled: `supplier_id`, `company_name`, `contact_name`, `contact_title`, and `city`. The table contains data for various suppliers, including Exotic Liquids, New Orleans Cajun Delights, Grandma Kelly's Homestead, Tokyo Traders, Cooperativa de Quesos 'Las Cabras', Mayumi's, Pavlova, Ltd., Specialty Biscuits, Ltd., PB Knäckebröd AB, and Refrescos Americanas LTDA.

	supplier_id	company_name	contact_name	contact_title	city
1	1	Exotic Liquids	Charlotte Cooper	Purchasing Manager	London
2	2	New Orleans Cajun Delights	Shelley Burke	Order Administrator	New Orleans
3	3	Grandma Kelly's Homestead	Regina Murphy	Sales Representative	Ann Arbor
4	4	Tokyo Traders	Yoshi Nagase	Marketing Manager	Tokyo
5	5	Cooperativa de Quesos 'Las Cabras'	Antonio del Valle Saavedra	Export Administrator	Oviedo
6	6	Mayumi's	Mayumi Ohno	Marketing Representative	Osaka
7	7	Pavlova, Ltd.	Ian Devling	Marketing Manager	Melbourne
8	8	Specialty Biscuits, Ltd.	Peter Wilson	Sales Representative	Manchester
9	9	PB Knäckebröd AB	Lars Peterson	Sales Agent	Göteborg
10	10	Refrescos Americanas LTDA	Carlos Diaz	Marketing Manager	Sao Paulo

Load Data

- Load data

MyNorthwind - Power Query Editor

File Home Transform Add Column View Tools Help

Close & Apply Close New Source Recent Sources Enter Data Data source settings Manage Parameters Refresh Preview Properties Advanced Editor Manage Choose Columns Remove Columns Keep Rows Remove Rows Sort

Queries [3]

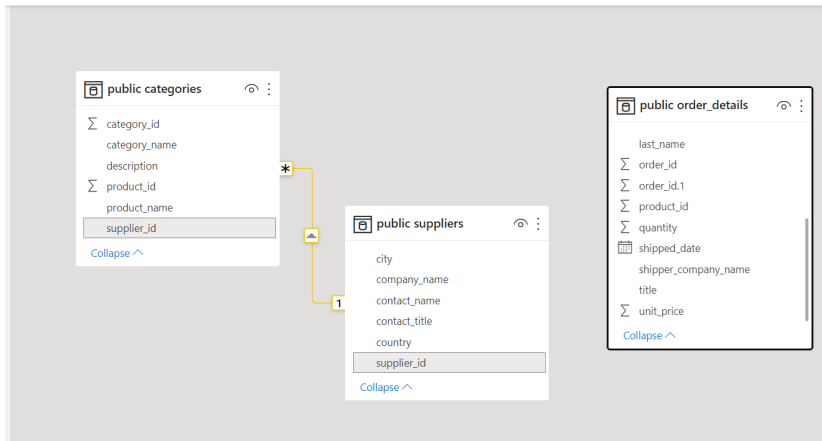
- public categories
- public order_details
- public suppliers

= Table.RemoveColumns("#Renamed Columns",{"public.products"})

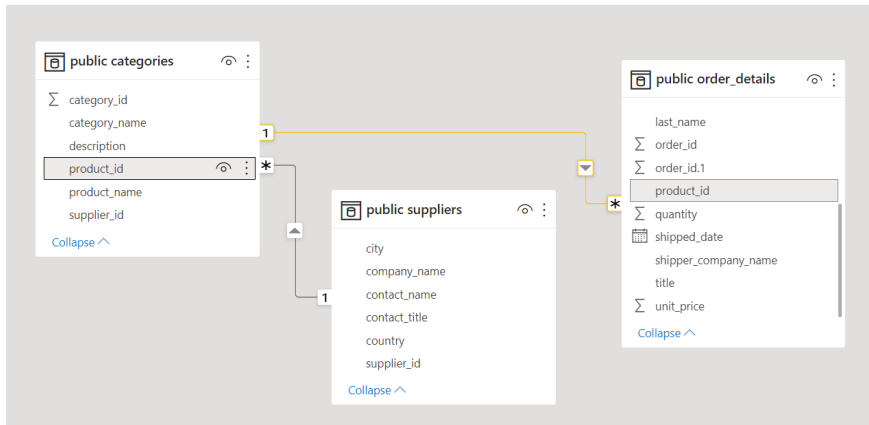
	1.2 order_id	1.2 product_id	1.2 unit_price	1.2 qua
1	10248	11	14	
2	10248	42	9.80000019	
3	10248	72	34.79999924	
4	10249	14	18.60000038	
5	10249	51	42.40000153	
6	10250	41	7.699999809	
7	10250	51	42.40000153	
8	10250	65	16.79999924	
9	10251	22	16.79999924	

Data model

- Correct the data model if necessary



Data model



Adding new columns and new measures

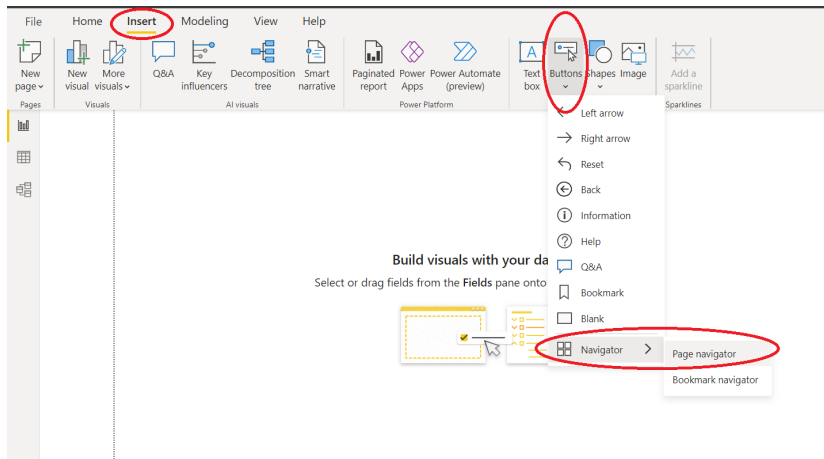
- `discount -> percent of discount %`
- New column Gross Revenue: `Gross Revenue = 'public order_details'[unit_price] * 'public order_details'[quantity]`
- New column Discount (\$): `Discount ($) = 'public order_details'[Gross Revenue] * 'public order_details'[discount %]`
- New column Net Revenue: `Net Revenue = 'public order_details'[Gross Revenue] - 'public order_details'[Discount ($)]`

Adding new columns and new measures

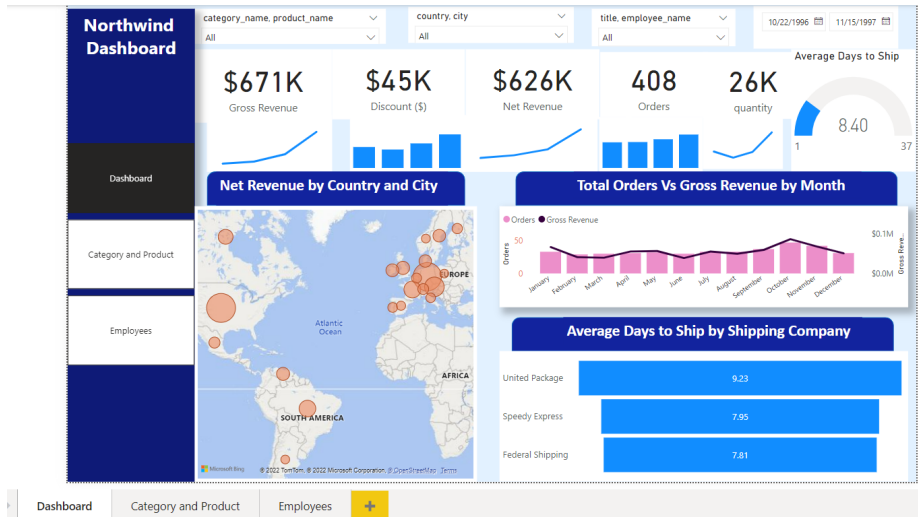
- New measure Orders: `Orders = DISTINCTCOUNT('public order_details'[order_id])`
- New measure Net Revenue per order: `Net Revenue per order = DIVIDE(SUM('public order_details'[Net Revenue]), 'public order_details'[Orders], BLANK())`

Visualization

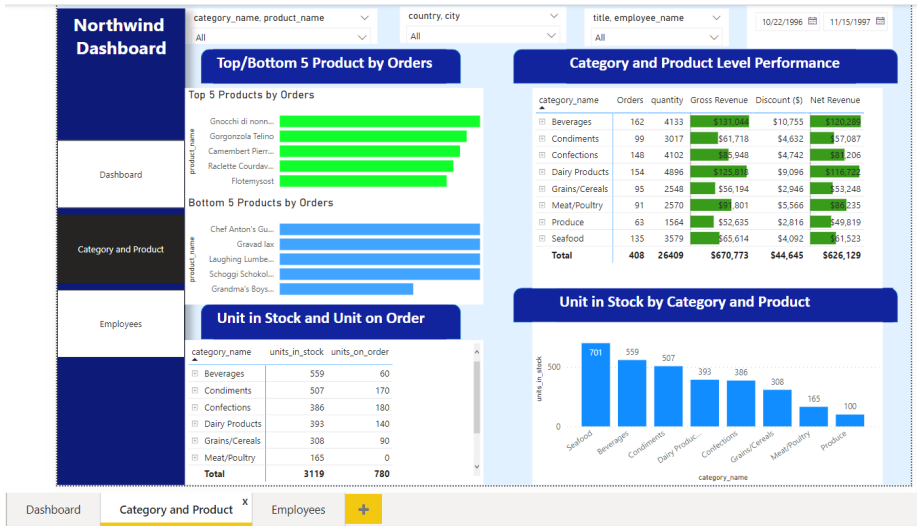
- Follow the video [Beautiful Power BI Dashboard with Northwind Database](https://www.youtube.com/watch?v=czON7fhEuYI&t=1599s) <https://www.youtube.com/watch?v=czON7fhEuYI&t=1599s> starting from 28:00. But
 - We do not use any background picture
 - We create a 3-page report with page navigators



Visualization



Visualization



Visualization



License



This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-nc-sa/4.0/).