29/6/24 Microsoft PL-300 Exam

→ Preparing the data:

↳ Query Editor: Used to load & transform our data. Mainly for cleaning the data before making visualizations & interpretations.

↳ Data Transformations:

* Shape & transform the table: change column names. By using "use first row as Headers" option.
* User-friendly name conventions.
* Resolving inconsistencies & Null values: Can be done using "Replace Values" tool.
* To remove duplicates:

  Select the unique column → Remove Rows → Remove duplicates.

* Evaluate & transform Column data types.
* Append queries: Only for same data structure. Adding rows.
* Merging queries: Different data structure. Adding columns based on 1 or more common columns. Works similar to joins in SQL.
* Pivot & Unpivot: Pivot: Rows into columns. (More readability).

  Unpivot: Columns into rows. (For reporting & analysis).

  Transform data → Pivot Column.
  → Unpivot Columns.

* Conditional Columns: Create new columns based on a condition using another columns in the dataset.

  Add Column → Conditional columns.

  (eg) Creating a Monthno column based on the monthname column based on a condition.

* Columns from example: This is used when we don't the conditions that must be implemented but know only the end result. This also adds additional columns.

  Add column → Columns from examples.

* Transpose: Process of flipping such that rows → columns & columns → rows. (For better Data Analysis & Visualization).

  Transform data → Transpose Data.

* Data Profiling: Must be done before loading the data for visualization.

  ↳ Column Distribution: describes the distribution of values in a column by a visualization (bar chart). It also gives the no. of distinct & unique values.

  (eg) null, null, 2, 2, 2, 4 ⇒ 3 distinct values (null, 2, 4)
  ⇒ 1 Unique value (4).

  ↳ Column Quality: describes valid, error, empty

  ↳ Column Profile: Distribution + Quality. Indepth statistics.

↳ Getting data from different sources:

* Files from folders: For this, all the files must have the same data structure inside the folder.
* Relational SQL DB: Get data → Connect to DB → Authenticat
  → Data Connectivity modes:
    * Import: Takes a local copy of the Data to the Power BI file & store it in-memory (cached).
    * DirectQuery: Possible for DBs. Data is not cached but retrieved on the fly directly from the DB.
    * Dual: Depending on the context, PBI determines what should be cached & what not.
  → Factors for deciding the connectivity mode:
    * Query Performance * Large dataset * Data Refresh.
* Dataverse: Lets us securely store & easily manage data that used by business applications. No-code Database.
* Data loading with parameters:
  Home → Manage Parameters.
  → Used to dynamically change the data source.

→ Model the data:

↳ Design a data Model:

* Data Model refers to the connection/ relationship b/w our
* Star Schema: Data type model.

→ Simplified models.
→ FT: observational or event data values: Sales orders.
→ DT: details of the data in the FT: weight of sold

Fact	Dimension
* Repeated values	* No repeated values
* For aggregation, visualization	* Not for aggregation but for grouping.
* Typically numbers and dates	

* Cardinality: Many to one (* : 1) → Foreign key to Primary k
  One to Many (1 : *) → PK to FK.
  Many to Many (* : *) → FK to FK.
  one to one (1 : 1) → PK to PK.
* Cross filter direction: Single, Many (Both).

* Common date table (DT):
→ We must always use a separate date dimension table. It can be:
* Imported from external resources (eg. Excel).
* Using DAX.
Data view → Table tools → New table.
DateTable = CALENDARAUTO() // Creates a complete list of all the dates in the fiscal year.
* using Power Query Editor
By utilising M-Query.

→ Data granularity: Refers to the level of detail or precision of the data stored in a dataset.
* High Granularity: Daily sales Data.
* Less Granularity: Monthly sales Data.

→ We can have only 1 active relationship b/w two tables.
→ We can use USERELATIONSHIP DAX function to use inactive relationships.

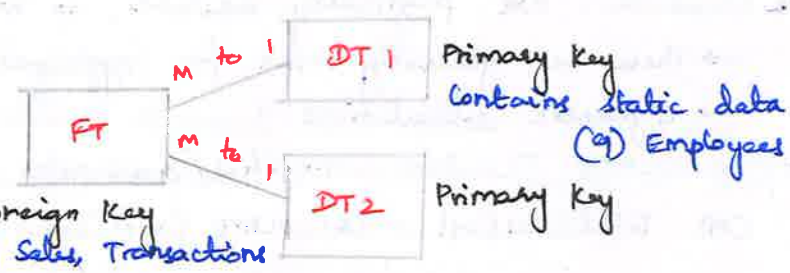
↳ Data Analysis Expression (DAX):
* Exciting in the form of Calculated Columns, tables, Measures.
* Calculated Columns:
→ Revenue Calculated = Sales [price] * Sales [sales]
→ CC can take up additional spaces as they are stored in the .pbix file.
* Best Method to avoid space consumption is to create columns in the following order:
Source Data → Query Editor → Calculated column.
* Measures: Not physically available, so no additional storage.
→ These are generally used for aggregations
→ Dynamic calculations.
* CALCULATE Function: Modifies/overwrites filter context of calculations.
(eg) TotalSalesEast = CALCULATE (SUM(Sales[SalesAmount]), Sales[Region] = "East") → Filter Context
// To calculate the total sales for a specific region.
→ We can also use multiple filters inside this function.
→ FILTER: A type of filter within calculate function.
→ ALL: Removes all the filters applied to any table/column.
* Time Intelligence: Specialized DAX functions used to manipulate data that has date & time dimensions.
(eg) DATEADD - Shifts the dates by a specified no. of intervals.
DATEYTD - Returns year-to-date for the current selection.
used for comparison over years/months/quarters.
* Semi-additive measure: Aggregate values just upto a certain point but not any further than that.

* Statistical functions: MIN, MEDIAN, MAX, VARIANCE, AVERAGE d
* Quick measures: choosing from a list of common calculations an add the results to the selected table. Built-in DAX calcul

↳ Optimize model performance: This depends on:
→ DAX measures
→ Small model size
→ Low level of cardinality. (Refers to the no. of distinct
(eg) 1, 2, 2, 3 ⟶ 3 distincts.
Low cardinality = better performance.

* We can ensure small model size + low cardinalities by usin the correct data types.
(eg) changing Date/Time to Date.
* Also, deleting unnecessary columns & rows.
* Avoiding repeated values.
* Replacing numeric values with measures. This reduces the data loaded into our data model.
* Summarizing data where possible.
* DAX measures: This can be modified by using Performance A
view ⟶ Performance Analyzer.
→ This gives the time for all our visualizations to loa
→ Based on the time, we can modify the DAX to ge better results & performance metrics.

QUIZ:
① Microsoft recommends splitting up date & time into separate col for better compression rates.
②



③ To add the full weekday name using DAX: DDDD.
④ Role-playing dimensions: Allows us to utilize the same dimension table in multiple contexts, providing flexibility & clarity in our DA

→ Visualize & Analyze the Data:
↳ Create reports:
* Table & matrix: Table for 2-D, Matrix for 3-D data.
* Column chart: For different categorical data.
* Line chart: Continous data: Used to determine the increasing/dec trends.
* Combo charts: Line & Column charts
* Tree maps: Clustered rectangle maps.

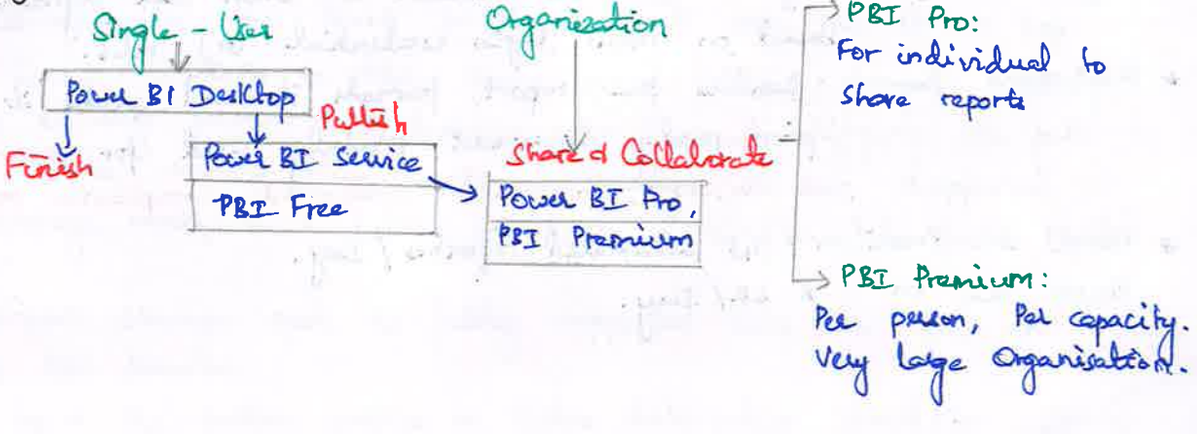
* cards: Used to display single value (such as cost, revenue).
* Gauge: Similar to cards but is used when we have a target data.
* Funnel visual: To show some progress.
* Waterfall chart: Breakdown of money (eg) Revenue, Profit, Sales etc.
* Customised tooltips: Tooltips can be created in a separate and later can be included to our main report page thus creating a customised tooltip visualization.

↳ Enhance reports for usability and storytelling:

* Sync Slicers: If we apply a filter on a slicer in page 1, it should also reflect on the Page 2. This is done using sync slicers.
* Drill throughs: Enable users to navigate from one report page to another, focusing on the selected data context.
* Drill downs: Allows users to explore data at different levels of granularity within the same visual. It is mainly used in bar / column charts.
* Bookmark: Stores a specific filter state of a page as a view.
  → Create a button.
  → Assign an action: Bookmark.
  → Enable bookmarks from view menu.
* Page navigation: Allows us to navigate to different report pages using buttons.
* Analyze Features: Built-in tools to analyze our data - in depth. (eg) Why there is a particular increase in trend, decrease in trend etc.
* Identify outliers: Scatter plot is generally used to visualize outliers.
* Forecast feature: Related to time-series analysis. Associated with line charts.
* AI features: Requires premium capacity.
  → Text Analytics
    * Sentiment Analysis
    * Key phrase extraction
    * Language detection.

→ Deploy & maintain Assets:

↳ Manage Workspaces:



* Sensitivity labels: Can be applied on our reports, dashboards, datasets, dataflows to guard our sensitive content against unauthorized data access. To have it:
  → we need PBI pro (or) premium license + edit permissions.
  → Azure Info. Protection Premium license.
* Creating workspace: To share & collaborate with others. Need a pro (or) premium license. My workspace is just our workspace and we can't use it to share our reports with others.
* Workspace roles:

Admin > Member > Contributor > Viewer

Highest Privilege — Add/remove users — Least privilege

Member: has the least privileges that can publish an app.

Cannot Add/remove users

↳ Create dashboards:

*	Reports	Dashboards
	→ For in-depth data analysis & exploration.	→ For quick, high-level monitoring & insights.
	→ Multi-page view	→ Single-page view.
	→ Extensive interactivity with the data	→ Limited interactivity.

* Live page: The dashboard reflects the report page. We can use filters in the live page. Any changes made in the report page is also reflected in the dashboard.
* Q & A feature: uses NLP to give suggestive visualizations for user's ques.
* Get insights: uses ML to give insights about our data to users.

↳ Manage files & datasets:

* Data Gateway: used to create a bridge btwn the powerbi cloud service & our local on-premises data center. It is a software. By using a data gateway, we can refresh our dataset in the PBI service.
* Scheduled refresh: To refresh our dataset automatically based on a refresh frequency.
* Role level security: usually done in PBI Desktop. Creates roles that enables constraints & restrictions.
  Static RLS
  (eg) Role – To see only the data in california.
* Dynamic RLS: Adjusts the data shown to each user dynamically based on their login credentials (or) roles.
* Paginated Reports: Renders the report, formats it and creates it as a print-ready document. Mainly used for invoices & receipts, financial statements, Compliance reports.
* Power BI Premium: 48 scheduled refreshes / Day.
  Power BI Pro: 8 SR / Day.

PRACTIS

① Drill
a so
are
② Dyna
for
③ It
and
④ Benef
⑤ To
⑥ only
⑦ Itera
colum
⑧ 2 m
⑨ A ba
categ
⑩ "ddd
"ddd
"dd"
⑪ Highes
to sh
⑫ Bookm
⑬ Steps
⑭ Hidde
⑮ Steps
⑯ CALCU
⑰ Permi
⑱ Data
supports
of the
⑲ the
views.
⑳ The
insights
㉑ The