

Excel BI-Training 70-779

Im Februar 2019 im Rahmen der Vorbereitung auf die Microsoft-Zertifizierungs-Prüfung "Analyzing and Visualizing Data with MS Excel" (70-779) erfolgreich durchgeführte Trainingsaufgaben gemäß vorgegebenem Hersteller-Leitfaden.

1.) Consume and transform data by using Microsoft Excel (Feature "Power Query")

1.1.) Import from data sources

- Benötigte Daten aus heterogenen Datenquellen und Systemen akquirieren:
- **CSV**: Parameter der Importspezifikation korrekt wählen (Separatoren, Encoding). AutoTypErkennung richtig nutzen.
- **JSON**: Excel 2016 native Workarounds a) via Textfilter und b) via Webfilter adressieren und in PowerQuery importieren.
- **Web**: Eine Ländertabelle aus Wikipedia-Seite importieren.
- **XML**: In Baumstruktur browsen und als mehrspaltige Tabelle importieren.
- **Excel**: Native Dateiverbindung zu anderer Exceldatei einrichten. a) Ganzes Blatt, b) nur benannten Bereich importieren. a) Wertkopieren importieren, b) nur Zellverknüpfung, c) als Excel-Abfrage-Objekt.
- **Verzeichnis**: Aus einem Verzeichnis 6 homogene Exceldateien als 1 Tabelle in PowerQuery importieren. Daten in Excel-Arbeitsblatt anzeigen lassen.
- **Microsoft Access**: In Access für die Kennwort-geschützte Beispieldatei "AdventureWorks2014.accdb" die Verschlüsselungsmethode ändern auf "legacy". In Excel eine neue Workbook-Verbindung zur Access-Quelltabelle "DimAccount" definieren.
- **Microsoft SQL Server**: Setup SQL Server 2017 Developer Edition, OLTP-Beispieldatenbank "Adventure Works 2016" (73 Tabellen, 492 Felder, 761k Datensätze). Von Excel aus Verbindung aufbauen. Tabellen Bestellungen (31k) und Bestellpositionen (121k) a) als Wertkopien in Arbeitsblatt importieren, b) nur Arbeitsmappen-Verbindung definieren, c) in PowerPivot-Datenmodell importieren, d) nur verbinden.
- **SSAS a) MultiDimensionales Modell**: Setup OLAP-Beispieldatenbank "Adventure Works Internet Sales". Von Excel aus Verbindung aufbauen und Daten aus Cube abfragen, Auswahl a) via Object Navigator, b) via MDX-Abfrage.
- **SSAS b) Tabellarisches Modell**: Setup OLAP-Beispieldatenbank "Adventure Works Internet Sales". Von Excel aus Verbindung aufbauen und Daten aus Modelltabellen abfragen. Über Verbindungsassistenten Verbindungsdefinition und Abfrage-Statement als odc-Datei speichern. Auch hier analog Auswahl via a) Object Navigator, b) via DAX-Abfrage.

1.2.) Perform data transformations

- In SQL Server die OLAP-SQL-Beispieldatenbank "Adventure Works Data Warehouse 2016" installieren (33 Tabellen, 365 Felder, 1 Mio. Datensätze).
- In Excel Verbindungen zur Faktentabelle "FactInternetSales" und die Dimensionstabelle "DimCustomer" definieren (ohne Import). In PowerQuery M-Abfrage nur auf die benötigten Spalten reduzieren. Spalte Discount% Datentyp von Dezimal nach Prozent ändern. Spaltenreihenfolge sinnvoller anordnen.
- Aus der Spalte Bestelldatum die neuen berechneten Spalten für Jahr, Monat, Monatsname extrahieren. Aus der Spalte Email die neue Spalte Username extrahieren.
- Die 3 Spalten Vorname, Mittelname und Nachname zu 1 neuen Spalte "Full Name" zusammenfassen.
- M-Abfragen organisieren in logische Ordner "Fact Queries" und "Dimension Queries".
- In PowerQuery das Ergebnis des "Query Folding" als optimierter nativer T-SQL-Code überprüfen.
- In PowerQuery 5 homogene Exceldateien in einem Verzeichnis in 1 M-Abfrage importieren und zu 1 Tabelle zusammenfassen. 6. abweichende Exceldatei mit separater M-Abfrage importieren, anpassen und der Zusammenfassung anfügen (Table.Combine()).
- Drei M-Abfragen zu den Exceldateien für ProductCategory, ProductSubcategory und Products erstellen. Key-Spaltentypen anpassen. Left-Join der ersten beiden Tabellen erstellen. Nested-Join und ExpandTableColumn anwenden. Ergebnis ist eine vollständig denormalisierte Tabelle.
- Data Cleansing: NULL-Werte durch default-Werte ersetzen. Duplikate entfernen.
- So entstandene M-Abfrage-Hierarchie als Diagramm anzeigen lassen.

- M-Abfrage erstellen, die aus obiger OLTP-Datenbank einen Server-seitigen Join aus den beiden Quelltabellen Customer und Person abfragt.
- In der Server-Tabelle "Customer Demographics" enthaltenes XML-Feld abfragen und in einzelne Spalten expandieren.
- Neue M-Abfrage mit Left-Join zwischen Customer und Customer Demographics anlegen. Da 2 verschiedene Quellen involviert sind, werden die 2 Sicherheitsstufen für die Überprüfung der Datenschutzebene abgefragt. 2x "Privat" setzen um beim Data Folding einen unkontrollierten Datenabfluß vollständig zu unterbinden.
- Neue M-Abfrage um die Textdatei "SalesTerritories" mit einer Hierarchie im Rollupformat zu importieren und in 3 separate Spalten zu zerlegen: Territory, Country, Continent (z.B. Northwest, United States, North America).
- In PowerQuery eigene UserFunction "ConvertCountryNameToISO-ALPHA-3" definieren. Mit Testmaske prüfen ("Canada" wird zu "CAN"). Funktion in M-Abfrage "Internet Sales All Countries" einbinden.
- Mit dem Feature "Bedingte Spalte" aus der Währungsspalte YearlyIncome eine neue Spalte "Income Category" mit 4 möglichen Testwerten erstellen.
- Neue M-Abfrage die den Umsatz nach Jahren aggregiert.
- M-Abfrage-Parameter für Jahr definieren. Mit Testmaske prüfen. Begrenzte Wertauswahl 2010 bis 2013 hinterlegen.

1.3.) Cleanse data

- Import der Tabelle CompetitorSales, deren Jahr-Spalte nur zu Beginn eines neuen Jahres befüllt ist, die restlichen Zeilen enthalten keine Jahresangabe. Auffüllen dieser leeren Zellen mit der "Ausfüllen nach unten"-Funktion.
- Bereinigen der Land-Spalte um die überflüssige Bemerkung " (look bob)" mit replace- und trim-Befehl.
- M-Abfrage erstellen für den Import eines formatierten Excelberichts mit mehreren, je Jahr partiellen pivotierten Umsatz-Tabellen. Schrittweise reduzieren auf die relevanten Datenbereiche. Daten entpivotisieren in die Spalten Jahr, Monat, Produktkategorie und Umsatz. Diese so bereinigten Daten schließlich wieder pivotisieren zu den Spalten Year, Month, Bikes, Accessories und Clothing.

2.) Model data (Feature "Power Pivot")

2.1.) Create and optimize data models

- In PowerPivot aus der oben erwähnten OLAP-SQL-Datenbank die 2 Tabellen DimProduct und FactInternetSales als Modelltabellen importieren (600 und 60k Zeilen).
- Anzeigen des ERM-Diagramms, in dem bereits die automatisch angelegte Relation zwischen beiden Modelltabellen angezeigt wird.
- Weiterhin die Tabellen DimProductCategory und DimProductSubcategory in das Modell importieren. Ergebnis: Auch hier wird die in der Quell-Datenbank definierte Relation zwischen diesen beiden neuen Tabellen in das Modell mitübernommen, nicht jedoch die Relation von SubCategory zu Product. Diese fehlende Relation manuell mit drag & drop definieren.
- Aus einem Beispiel-Arbeitsblatt 2 Table-Objekte in das Datenmodell übernehmen als "Linked Tables". Automatisches Update aktivieren. In Arbeitsblatt neuen Datensatz anfügen. Ergebnis Aktualisierung in PowerPivot-Fenster überprüfen.
- Neue Exceldatei anlegen. In das Modell die Quelltabellen DimDate und FactInternetSales importieren (3,6k und 60k Zeilen). Ergebnis im ERM-Diagramm sind 3 automatisch mitübernommene Relationen, von denen jedoch nur 1 aktiv ist.
- In Beispiel-Arbeitsblatt mit 2 Table-Objekten Person und Umsatz, zwischen denen noch keine Relation definiert ist, Pivot-Tabelle in Arbeitsblatt einfügen. Ergebnis sind falsche Summen aufgrund der fehlenden Relation. Das Excel-Feature "Automatische Erkennung" (von Beziehungen) aufrufen. Ergebnis ist das automatische Anlegen der fehlenden Relation, und in Folge korrekte Summen in der Pivot-Tabelle.
- Optimieren der Benutzbarkeit der Pivot-Tabelle durch Ausblenden nicht benötigter bzw. verwirrender Felder wie z.B. technische ID-Felder. Hinterlegen von Feldbeschreibungen. Definieren von verständlicheren Synonymen statt der technischen Tabellennamen (z.B. "Sales Person" statt "tblSalesPerson"). Anlegen von Perspektiven für verschiedene Zielgruppen mit reduzierter Feldauswahl je Perspektive. Wo sinnvoll Sortieren nach Fremdspalten hinterlegen (z.B. Monatsname nicht alphabetisch, sondern immer numerisch)

sortieren). Für die Benutzer zusätzliche praxisnahe Hilfsfelder bereitstellen (z.B. 3-stellige Wochentagsabkürzung).

- Kalendertabelle: a) Bestehende Modelltabelle als Kalendertabelle markieren oder b) neu erzeugen lassen mit einem Zeithorizont entsprechend allen in den Modelldaten vorhandenen Zeitdaten.

2.2.) Create calculated columns, measures, and tables (DAX)

- In neuer Exceldatei in Modell die Quelltable FactInternetSales importieren. Im Arbeitsblatt eine Pivottabelle einfügen. Im Modell ein "explizites" Measure "Total Sales" mit der DAX-Formel für Summe(Sales) definieren. Zum Vergleich in der Pivottabelle mit drag & drop das "implizite" Measure "Summe von Sales" definieren. Ergebnis sind identische Summen. In PowerPivot Ergebnis implizites Measure anzeigen lassen.
- Verbindung zu tabularem SSAS herstellen, Tabelle mit DAX-Abfrage EVALUATE('Internet Sales') importieren. Datenvolumen sinnvoll begrenzen mit ORDER-BY-START-AT-Klausel.
- Weitere Performancesteigerung durch bereits Server-seitige Aggregation erreichen mit SUMMARIZE-Funktion. Zusätzliche Zwischensummen-Zeilen anfordern mit ROLLUP-Funktion. Measure-Definition von PowerPivot-Modell verlagern in DAX-Abfrage mit DEFINE-MEASURE-Klausel. Filtern nach z.B. Farbe = rot oder blau mit RELATED-, FILTER- und OR-Funktion.
- In Arbeitsblatt auf Modelldaten basierende Pivottabelle einfügen. Pivottabelle mit Feature "In (Cube-) Formeln konvertieren" umwandeln. Ergebnis ist der gleiche Zellbereich mit Cube-Formeln belegt (CUBEWERT). Beliebige Verschieben einzelner Zellen ändert nicht deren Wert. Das Berichtslayout kann also beliebig angepaßt werden.

2.3.) Create Hierarchies

- Im Modell in der Tabelle DimDate eine 4-stufige Hierarchie aufbauen und benutzerfreundlich umbenennen in: Calendar Year, Calendar Quarter, Month Name und Date.
- In einem neuen Arbeitsblatt eine neue Pivottabelle einfügen. Als Spalten die Produktkategorie, als Wert die Summe Umsatz und als Zeilen die gerade definierte Zeithierarchie wählen. Ergebnis sind Zeilenköpfe, die mit einem Plus-Symbol versehen sind und so einen Drilldown in der Zeitdimension bis auf die Ebene der einzelnen Tage ermöglicht.
- Die Monatsnamen werden alphabetisch sortiert. Eine Sortierung über das Fremdfeld MonthNumberOfYear hinterlegen.
- In der Tabelle DimProduct eine 3-stufige Hierarchie erstellen mit den Ebenen ProductCategory, ProductSubcategory und EnglishProductName. Hierzu die erforderlichen 2 Fremdfelder mit RELATED aus 2 Fremdtabellen DimProductCategory und DimProductSubcategory beziehen. Um die Feldauswahl benutzerfreundlicher und übersichtlicher zu gestalten, die beiden Fremdtabellen für den Benutzer ausblenden.
- Die gerade erstellte Produkthierarchie für die Spaltenbelegung wählen. Ergebnis sind jetzt auch in den Spaltenköpfen Plus-Symbole für den Drilldown in der Produkthierarchie bis auf die Ebene der Einzelproduktnamen.

2.4.) Create Performance KPIs

- Im Modell Measure "SalesPY" definieren, das den Umsatz im Vorjahreszeitraum berechnet. Ein weiteres Measure "SalesGrowth" anlegen, daß die Relation zwischen Jahresumsatz und Vorjahresumsatz berechnet.
- SalesGrowth als (Ist-)Basiswert für neuen Key Performance Indicator verwenden. Statusschwellenwerte für die Colorierung der Ampelfarben bei dezimal +/- 0,05 (= +/- 5%) setzen.
- In einem leeren Arbeitsblatt eine neue Pivottabelle erstellen und als Spalten wählen: Sales, SalesPY, SalesGrowth Status und SalesGrowth. Ergebnis z.B. -3,2 % wird noch als gelb angezeigt.

3.) Visualize data

3.1.) Create and manage PivotTables

- Komplexes Modell aufbauen mit FactInternetSales und 6 Dimensionstabellen. Die Benutzerfreundlichkeit verbessern: Alle technischen Key-Felder für den Benutzer ausblenden. In PowerQuery den Prefix "Dim" 6x

entfernen. Für die Datumswerte die Uhrzeit " 00:00:00" ausblenden über die Feldformatierung in PowerPivot.

- Im Excel-Ribbon "PivotTable Entwurf" das Aussehen optimieren: Eines der vordefinierten Farbschemata wählen für besseren Kontrast, überflüssige Teilergebnis-Zeilen und Spalten ausblenden.
- Die Pivottable um 3 interaktive Filterelemente erweitern:
- Pivottabellen-Seitenfilter mit Feld "PromotionKey" einrichten.
- Ein Datenschnitt-Steurelement ("Slicer") auf Basis des Gender-Feldes erstellen. Der Benutzer kann so durch einen Klick auf einen Auswahlwert in der Liste "F", "M" schnell die Pivottable filtern, ohne erst einen Seitenfilter einrichten zu müssen.
- Ein Zeitachsen-Steurelement ("Double Slider") auf Basis des OrderDate-Feldes erstellen. Der Benutzer kann hier mit 2 Schiebereglern schnell Anfang und Ende des Filterzeitraums setzen. Hierbei kann der Benutzer zwischen den Zeiteinheiten Jahre, Quartale, Monate und Tage wählen.

3.2.) Create and manage PivotCharts

- a) 7+1 Excel-Diagrammtypen, die ein Excel-Pivottabellen-Objekt als Datenbasis unterstützen: 2D-Säule, Linie, Kreis, Balken, Fläche, Oberfläche, Netz und Verbund.
- Im Begleitmaterial vorhandene Demodatei 3-3 erweitern um neues Arbeitsblatt. Pivottable neu anlegen mit den Zeilen SalesTerritoryGroup, SalesTerritoryCountry und der 1 Measure-Spalte Total Sales. Auf Basis dieser Pivottable 2 Diagramme vom Typ Säule und Linie erstellen.
- Auf gleicher Basis Verbund aus Säulen- und Linien-Diagramm erstellen.
- b) 10 Diagrammtypen, die ein Excel-Pivottabellen-Objekt nicht(!) als Datenbasis unterstützen: 2D-Punkte, Kurs, TreeMap, SunBurst, Histogramm, Kastengrafik, Wasserfall, 3D-Karte, Trichter und 2D-(Welt-)Karte.
- Hier Workaround anwenden: Neues Arbeitsblatt erstellen, Pivottable wie oben anlegen. Diese dann in Cube-Funktionen konvertieren. Anschließend diesen Zellbereich als Datenbasis verwenden für neue Diagramme vom Typ TreeMap und SunBurst.