

Checkliste: Best Practices Business Intelligence (mit Excel oder Power BI)

Für die Umsetzung eines Business Intelligence Projekts haben sich einige Aspekte als besonders wichtig herausgestellt: sogenannte „Best Practices“. Diese Best Practices werden in diesem Dokument am Beispiel von Power BI Desktop und Excel (Power Query / Power Pivot) praxisnah erläutert. Diese beiden Tools bieten sich hierfür besonders an, da sie für die Mehrheit der SwissAccounting Mitglieder frei verfügbar sind.

Das vorliegende Dokument ist keine Schritt-für-Schritt Anleitung und setzt bereits Anwenderkenntnisse in den obengenannten Tools voraus. Entsprechende Kurse werden von SwissAccounting, der Controller Akademie in Zürich und weiteren Lehrinstitutionen angeboten.

Das Vorgehen bei einem Business Intelligence Projekt kann vereinfacht in drei Teilschritten dargestellt werden (siehe Abbildung unten).

Datenquellen

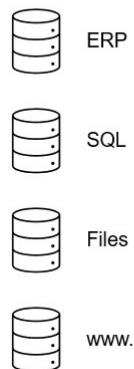


Abbildung 1 - Vorgehen bei Business Intelligence Projekt

Zum letzten Schritt - der «Visualisierung» - gehört zudem das Berechtigungsmanagement sowie verschiedene Arten der Berichtspräsentation. Für jeden Schritt können, je nach Unternehmenskonstellation, verschiedene Tools eingesetzt werden. Manche Tools integrieren alle diese Funktionen in einem, wie bspw. Power BI.

1. Datenvorbereitung (ETL)

Checkpoint	Beschrieb	Beispiel	Pro Tipp
<input type="checkbox"/> Keine Datenfelder auf Vorrat	<p>In ein Datenmodell werden nicht alle Rohdaten geladen, sondern nur tatsächlich benötigte Daten. Entsprechend ist im ETL-Prozess eine Auswahl vorzunehmen. Dies erhöht die Performance und reduziert die Komplexität.</p>	<p>In das BI der Hotelkette Goodsleep AG wird die Abacus Tabelle mit allen Buchungen eingebunden. Diese enthält 180 Spalten. Im ETL-Prozess werden nur 10 bis 20 Spalten ausgewählt (bspw. Kostenstelle, Konto, Betrag, Buchungsdatum) und in das Datenmodell geladen.</p>	<p>Im Power Query Editor ermöglicht die Funktion <i>Andere Spalten entfernen</i> die Auswahl der Spalten, welche beibehalten werden sollen. Dies führt in der späteren Anpassung/Automation zu weniger Problemen als die Funktion <i>Spalten entfernen</i>.</p>
<input type="checkbox"/> Für alle Spalten sind Datentypen definiert	<p>Allen Spalten muss ein Datentyp (bspw. Dezimalzahl, Text, Datum) zugewiesen sein. Falsche oder nicht zugewiesene Datentypen können zu Auswertungsfehlern führen.</p>	<p>Im Gastronomiebereich der Goodsleep AG wurde bisher nur der Umsatz auf Tagesebene betrachtet. Neu soll der Umsatz im Tageszeitverlauf analysiert werden, um zu klären, ob sich das Nachmittagsgeschäft rentiert. Das Datumsfeld, welches in den Rohdaten auch die Uhrzeit beinhaltet, muss neu auf den Typ "Datum/Uhrzeit" umgestellt werden.</p>	<p>Die sogenannten „Schlüssel“ zwischen den Tabellen (z.B. Mitarbeiternummern, Projektnummern, Kontonummern etc.) werden idealerweise als Text formatiert, da mit diesen Nummern – anders als bei Finanzkennzahlen – nicht gerechnet werden soll.</p>

<input type="checkbox"/> Die Daten- transformation ist verallgemeinert / automatisierbar	<p>Der ETL-Prozess ist immer als Automation mit generalisierten Regeln umzusetzen. Das bedeutet, dass die konfigurierten Datentransformationen auf neuen Daten (z.B. aus dem Folgejahr) genauso anwendbar sein müssen.</p>	<p>Die Kundenumfragen der Hotelgäste der Goodsleep AG kommen via Excel ins BI und haben eine Spalte mit dem aktuellen Jahr. Die Spaltenüberschrift sollte somit das übergeordnet Dimensionselement "Datum" beinhalten anstelle bspw. «2025».</p>	<p>Laden Sie den ganzen Power Query Code in ein Word (in Power BI z.B. Datei --> Info --> kopieren) schauen Sie, ob dort konkrete Datumswerte oder Produktnamen etc. vorkommen. Dies sollte nicht der Fall sein.</p>
<input type="checkbox"/> Zuerst filtern, dann transformieren	<p>Oft werden nicht alle Einträge (Zeilen) benötigt, sondern bspw. nur alle nach 2020 oder alle Einträge ohne Storno-Kennzeichen. Das Filtern auf benötigte Einträge sollte am Anfang der Abfrage erfolgen, dies erhöht die Performance.</p>	<p>Im ETL-Prozess der traditionsreichen Goodsleep AG mit einer Datenhistorie von 25 Jahren wird als erster Schritt ein Datumsfilter (bspw. > 2015) gesetzt. Anschliessend werden alle anderen Transformationsschritte (berechnete Spalten, Zusammenführung von Tabellen etc.) durchgeführt.</p>	<p>Die nächste Ausbaustufe der Performanceoptimierung ist „Query Folding“. Dies ermöglicht, dass Abfragen resp. die ersten Schritte davon direkt in der Datenbank des Quellsystems durchgeführt werden. Online gibt es ausführliche Erläuterung dazu (bspw. Microsoft Learn).</p>
<input type="checkbox"/> Die Daten wurden im ETL- Prozess genügend bereinigt	<p>Viele Daten sind in Rohform (direkt aus dem Quellsystem) fehlerbehaftet oder inkonsistent und müssen im ETL-Prozess bereinigt werden.</p>	<p>Im Reservationssystem des Hotels werden regelmässig Testbuchungen abgesetzt, um das System zu testen. Diese Buchungen werden speziell gekennzeichnet und werden auch im ETL-Prozess fürs BI Vorhaben bereinigt, also herausgefiltert.</p>	<p>Die Funktion <i>Fehler ersetzen</i> in Power Query Editor bringen Stabilität in den g-Prozess. Langfristig lohnt es sich jedoch - je nach Fall - die Fehler in einer separaten Abfrage sichtbar zu machen und die Quellsysteme so anzupassen, dass diese Fehler in der Zukunft vermieden werden.</p>

2. Datenmodellierung

Checkpoint	Beschrieb	Beispiel	Pro Tipp
<input type="checkbox"/> Es wurde ein Star-Schema angewandt	Star Modelle gelten als Best Practice bei der Datenmodellierung, da sie die Übersicht gewährleisten und die Erweiterbarkeit des Datenmodells um weitere Tabellen problemlos möglich ist.	Goodsleep hat für ihren Umsatz die Datentabellen so vorbereitet, dass Sie zu einem Starmodell verbunden werden kann: Drei Faktentabellen (Umsatz, Budget und Zimmerstatistiken), sowie fünf Dimensionstabellen (Datum, Betrieb/Standort, Abteilung, Leistungsart und Segment).	Ordnet man entgegen dem Bild eines Sterns alle Dimensionstabellen oben und alle Faktentabellen unten an, kann die Beziehung einer Dimensionstabelle zu einer oder mehreren Faktentabellen besser erkannt werden. Die Beziehungen zwischen Dimensions- und Faktentabellen ist als one-to-many (1:n) zu konfigurieren.
<input type="checkbox"/> Keine beidseitigen Beziehungen oder Many-to-many Beziehungen (resp. nur in begründeten Ausnahmefällen)	Beziehungen mit beidseitigen Filterungen und many-to-many Beziehungen (*:*) können langfristig zu Problemen führen und sollten nur in begründeten Ausnahmefällen verwendet werden.	Im Power BI der Goodsleep AG wurde die automatische Beziehungserkennung übernommen und die Werte waren nicht mehr nachvollziehbar. Die Bereinigung der Beziehungen führte zu einem funktionierenden und um weitere Tabellen ausbaubaren Modell.	Power BI erstellt nach dem Laden der Daten eigenständig Beziehungen, diese Option kann deaktiviert werden. In jedem Fall muss bei jeder Änderung in der Tabellenstruktur das Datenmodell und die Beziehungen aktiv überwacht werden.

<input type="checkbox"/> Verwendung einer eigenen Datums-tabelle	<p>Im Grossteil aller Datenmodelle gibt es einen Zeitbezug. Soll ein Datenmodell erweiterbar sein und will man von den vorgefertigten Time Intelligence Funktionen in DAX profitieren, benötigt es eine separate Datumstabelle.</p>	<p>Die Hotelkette Goodsleep kann die eigene Datumstabelle als Berichtsfilter verwenden. Durch die Verbindung im Modell werden alle Faktentabellen (Umsatz, Budget und Zimmerstatistiken) parallel auf das gleiche Datum gefiltert. Ausserdem können einfache DAX-Funktionen wie «Year-to-date» verwendet werden.</p>	<p>Mit Power Query und einer entsprechenden Funktion kann die Datumstabelle so aufgebaut werden, dass sie dynamisch auch in Zukunft alle relevanten Datumswerte generiert.</p>
<input type="checkbox"/> Verwendung von expliziten Measures in den Berichten	<p>Ein Datenfeld wie Umsatz könnte direkt in einem Visual dargestellt werden und wird – basierend auf dem Datentyp – automatisch aufsummiert (s.g. «Implizites Measure»). In fortgeschrittenen Szenarien kann dies jedoch zu Problemen führen. Daher sollte jede Kennzahl explizit mit DAX definiert werden.</p>	<p>Die Goodsleep AG hat zunächst die Anzahl Übernachtungen als Datenfeld in ihren Visualisierungen verwendet (Implizites Measure). Nach einem halben Jahr kommen erstmals Berechnungsgruppen zum Einsatz, wodurch die Impliziten Measure nicht mehr funktionieren. Mit der DAX Funktion SUM() wurde ersatzweise ein explizites Measure definiert.</p>	<p>Hat man die Grundlagen von DAX verstanden, sind KI-Chatbots wie ChatGPT oder Copilot hilfreich bei der Erstellung und Fehlerbehebung von DAX-Code.</p>
<input type="checkbox"/> DAX Measures bauen aufeinander auf	<p>Hier gilt das DRY-Prinzip (Don't Repeat Yourself). Zuerst erstellt man ein Basis-Measure und darauf aufbauend spezielle Unterversionen, welche das Basis-Measure wiederverwenden. Langfristig werden so Inkonsistenzen vermieden.</p>	<p>Die Goodsleep AG hat ein Measure «EBIT» und ein Measure "adjusted EBIT" definiert. Der DAX-Code für das Measure «adjusted EBIT» bezieht sich dabei auf das «EBIT» Measure und filtert dann mit der "CALCULATE"-Funktion die Sondereffekte raus.</p>	<p>In einer leeren Tabelle «Measures» können die Measures abgelegt werden. Mit Ordnern unterhalb der Tabelle können dann Basis-Measures und spezifische Measures unterteilt werden.</p>

3. Visualisierung

Checkpunkt	Beschrieb	Beispiel	Pro Tipp
<input type="checkbox"/> Berichte sind auf die Zielgruppe ausgerichtet (inkl. Berechtigungen)	Die Berichte wurden zielgruppengerecht gestaltet und vor der ersten Veröffentlichung wurden die Berechtigungen getestet.	Die Goodsleep AG hat für die Organisation ein Berichtspaket erstellt, welches zwei verschiedene Versionen enthält. Ein grafisches Dashboard mit vielen aggregierten Werten für die Geschäftsleitung, sowie eine Version für Betriebsleiter mit einer Matrix in der alle Details zu ihrem täglichen Betrieb aufgeführt sind.	Mit der «Row Level Security» können die Zugriffsberechtigungen sehr granular gesteuert werden. Somit kann ein Bericht für verschiedene Zielgruppen individuell gefiltert werden, statt dass mehrere verschiedene Berichte erstellt werden müssen.
<input type="checkbox"/> Daten sind mit dem Vorsystem abgestimmt	Daten aus dem Bericht wurden mit den Quellsystemen abgestimmt und sind – je nach Messgenauigkeit – "ausreichend" korrekt.	Die Goodsleep AG hat in der Datenbank immer wieder Probleme mit Buchungen, die von Plattformen wie Booking.com kommen. Diese werden nicht richtig in die Anzahl Logiernächte übernommen und alle Folgekennzahlen sind verfälscht. Eine manuelle Abstimmung legt diesen Umstand offen und es wird eine Lösung mit dem externen Datenlieferanten erarbeitet.	Erstellen Sie nicht nur ein inhaltliches Dashboard, sondern auch ein Datenqualitätsdashboard. In diesem Qualitätsdashboard werden die fehlerhaften Datenaspekte visualisiert (Dubletten, Extremwerte, unvollständige Datensätze etc.).

<input type="checkbox"/> Kennzahlen sind mit Kontext versehen	<p>Die meisten Zahlen sind ohne Kontext wenig aussagekräftig. Für alle Kennzahlen sollte daher Kontext in Form vom Budget, Vorjahreswerten, Benchmarks etc. gegeben werden.</p>	<p>Im Dashboard der Goodsleep AG ist auf der ersten Seite die Kennzahl "Anzahl Logiernächte" platziert. Diese wird in Rot dargestellt, sofern sie unter dem Vorjahreswert liegt. Die Abweichung von -5% wird ebenfalls direkt auf dem Dashboard dargestellt.</p>	<p>Viele Visualisierungselemente haben bereits Felder für den Kontext reserviert, wie z.B. im Power BI das "KPI" Visual.</p>
<input type="checkbox"/> Performance wurde getestet	<p>Heute ist die Erwartung von Berichtlesern, dass Berichte direkt ohne Warte- oder Ladezeit erscheinen. Dies muss vor dem Ausliefern getestet werden.</p>	<p>Das Management der Goodsleep AG sichtet jeden Morgen im Browser die Logiernächte des Vortages im Bericht. Die Seite hat genauso schnell geladen wie ein Newportal oder eine andere Webseite.</p>	<p>Mit der Funktion «Performance Analyzer» in Power BI Desktop kann die Ladezeit eines Reports getestet und verlangsamende Faktoren identifiziert werden.</p>
<input type="checkbox"/> Bericht ist übersichtlich und die Bezeichnungen klar	<p>Bei einem Dashboard gilt die Analogie zu einem Cockpit im Auto: Sind für die Benutzer die relevanten Kennzahlen in wenigen Sekunden ersichtlich oder müssen sie diese lange studieren oder gar filtern? Die gezeigten Informationen sollten zudem unmissverständlich bezeichnet sein.</p>	<p>Die Goodsleep AG halbiert die Anzahl der Visuals auf ihrem Dashboard von 16 auf 8 und zeigt die Verkäufe statt pro Land neu pro Hauptregion.</p>	<p>Farben sind wichtig und jede Farbe signalisiert etwas anderes. Farben sollten daher sparsam eingesetzt werden und Signalfarben wie bspw. rot sollten für entsprechende Anmerkungen reserviert werden.</p>