



## Anfertigen eines CAT bzw. PIKE

**Autor: Andreas Schau**

### Kann ... // Algorithmus x // exact ... // (Kunden)-Problem ... berechnen / lösen?

Ja, mit Microsoft Power BI kann man sehr einfach und effizient Visualisierungen für große Datenmengen, zum Beispiel Verkaufsdaten aus verschiedenen Quellen, erstellen und diese dann in Reports und Dashboards verwenden. Es erlaubt außerdem eine einfache Daten-Aufbereitung und Bereinigung sowie die Erstellung von `berechneten Parametern` mit Hilfe der DAX-Programmiersprache. Die so gewonnenen Reports und Dashboards erlauben dem Kunden dann intuitive Einblicke in seine Daten, wodurch Fehler erkannt und neue Verkaufs-Ziele formuliert werden können.

### Data Science Kernaussage:

#### (P) roblem):

**[Welcher Frage hat für die Lösung des Kunden / Auftraggeber die größte Bedeutung?]:**

Befindet sich unser Unternehmen auf Erfolgskurs? (Fahren wir Verluste ein, die vermeidbar sind?; Welche Kundensegmente, Produktkategorien und Regionen sind am profitabelsten und bieten Expansionspotenzial?)

#### (I) ntervention:

**(Bibliotheken- und Algorithmen-Auswahl, ... z.B. pandas für Finanzdaten ... )**

**[Welche Berechnung erwäge ich vornehmlich?]:**

Excel und CSV Data-Loader für das Daten-Importieren, Power BI für die Datenaufbereitung, Filterung, Visualisierung und Report sowie Dashboard Erstellung.

#### (K) ontrollintervention

**(falls erforderlich: Bibliotheken- und Algorithmen-Auswahl ... z.B. scikit-learn für Finanzdaten ... )**

**[Was ist die andere Möglichkeit?]:**

Python bietet mit entsprechenden Bibliotheken ebenfalls alle Werkzeuge für die Datenvorverarbeitung, Analyse und Darstellung. Der Ersatz von Power BI durch ein Python-Script oder KNIME-Workflow ist möglich.

#### (E) rgebnismaß (Zielgröße(n)) – Die Evidence

**[Was möchte ich / der Kunde erreichen? Z.B. Prädiktor oder Klassifikator erstellen ... ]:**

Verluste Identifizieren: Darunter zum Beispiel Produktkategorien, einzelne Produkte, Kundensegmente oder Regionen herausstellen, die besonders schlecht abschneiden und dringender Intervention benötigen.

Gewinne Identifizieren: Ebenfalls nach diesen Dimensionen untersuchen um Top-Performer zu finden, die vor zu experimentellen Änderungen geschützt werden sollten.

Weitere Nützliche Informationen herausarbeiten, wie Zukunftstrends, Bearbeitungszeiten oder regional unterschiedliches Kundenverhalten.

### Anmerkungen

### Literaturhinweise

## Die Suche nach der besten Evidenz

### 1. Problem

Das Superstore Dataset ist sehr umfangreich, hat jedoch hohe Qualität. Deswegen müssen nur wenige Vorbereitungs-Schritte ausgeführt werden. Header korrekt setzen, "null"-Werte entfernen, Datentypen korrigieren und überprüfen, ob die Beziehungen zwischen den Tabellen korrekt erkannt wurden. Dann können die Daten explorativ analysiert werden. Dabei ist das Ziel Ausreißer zu finden, die dem Unternehmen schaden, sowie Top-Performer zu finden, damit diese geschützt und deren Erfolg auf andere Produkte, Regionen oder Mitarbeiter übertragen werden können.

## 2. Definition einer wichtigen suchbaren Frage

Welche Produkte, Produktkategorien, Standorte oder Mitarbeiter fahren Gewinne und welche Verluste ein?

## 3. Auswahl der wahrscheinlichsten Quelle für diese Evidenz

Visualisierungen, die diese Parameter mit dem Verkaufsvolumen und Profit, sowie Discount und auch den Retouren gegenüberstellen.

## 4. Erstellung einer Suchstrategie

Einlesen des Datensatzes samt Vorverarbeitung, Erstellung mehrerer relevanter Visualisierungen und Trend-Darstellungen. Beginnend bei groben Kategorien, bis hin zu sehr detaillierten Kategorisierungen und sogar einzelnen Produkten und Bestellungen, wenn nötig.

## 5.0 Zusammenstellung der Evidenzausbeute

Zusammenfassung der anschaulichsten Visualisierungen in einem Dashboard.

- „Summe von Profit nach State und Region“: Zeigt über Kreisgröße welche Regionen den meisten Gewinn einfahren.
- „Summe von Profit nach Sub-Cateogry und Segment“: Zeigt, dass „Tables“ in allen Segmenten Verluste einfahren, sowie, dass „Bookcases“ und „Supplies“ im Konsumer-Segment ebenfalls Verluste bringen.
- „Erstes Datum: Product Name nach Segment, Profit und Discount“: Zeigt, dass einige 3D-Drucker nur mit Discount zu starken Verlusten verkauft werden konnten. Filtert man, via klick in einem anderen Diagramm, nach Tischen, so sieht man diese ebenfalls mit hohen Discounts große Verluste einfahren.
- „Summe von Profit nach Jahr, Quartal, Monat und Category“: Zeigt trotzdem einen positiven Profit-Trend aller Kategorien über die letzten vier Jahre.

## 6. Anwendung der Evidenz

Die Geschäfte sollten ihr Feingefühl für 3D Drucker und Tische weiter ausbauen oder diese Produkte anderen Wettbewerbern überlassen. Da sie jedoch erstaunlich gut im normalen Drucker-Verkauf sind, können sie vielleicht in den nächsten Jahren ihr Know-How noch besser auf das neue Marktsegment des 3D Druckens anwenden.