

KNIME

Erstellt von

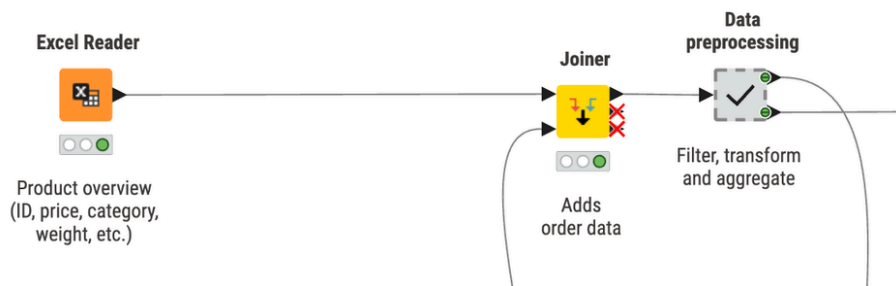
Petra Palinkas

03.03.2026

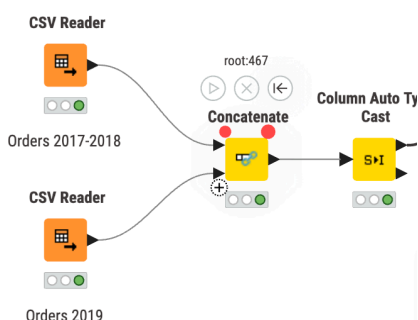
KNIME Projekt

Optimierung von Preisen in der Wirtschaft

Als erstes wurde zum Auslesen von Daten der Excel Reader verwendet. Damit konnte ich die Daten aus einer heruntergeladenen Datei von Preisen in den folgenden Kategorien auslesen lassen.



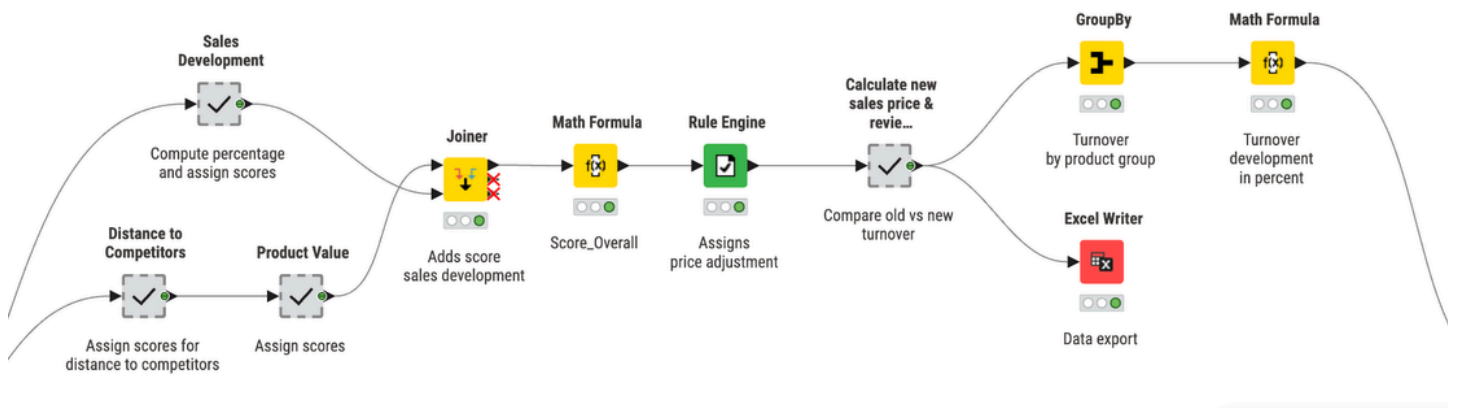
ID, Price, Category, Weight, etc. Danach hab ich mit einem Joiner andere Daten von zwei csv readern hinzugefügt, in welchem Alte Bestellung zu finden waren. Die wurden aber davor mithilfe von einem Concatenate (Dieses Node verkettet zwei oder mehrere tables miteinander). Danach wurden die Datentypen in die richtige Typen geändert. (Integer, String, usw. hierbei wird einfach angeschaut ob es zb ein Datum ist oder nicht und welcher Datentyp dazu passt)



Nach dem das erledigt war und wie vorher schon erwähnt beim Joiner hinzugefügt wurde, hat ein Data Preprocessing Node das ganze verarbeitet. Dieser Node ist schon für sowas vorhergesehen und hat mehrere Funktionen wie, data access von product Informationen und Bestellung zu verarbeiten.

Dies führte zu einem erneuten Joiner welcher aber auch mit einem weiteren Excel Reader, Daten von sozusagen Konkurrenten ausgelesen hat und das wieder mit einander zusammengefügt.

Danach führte das ganze weiter mit zwei Komplexen Teilen. Einerseits die obere:



Bei welchem als erster durch die verschiedenen Nodes die schon dafür gedacht waren, Sales Development, Distanz zu Konkurrenten und Produkt Wert ermittelt wurden. Die beiden wurden dann zusammengefügt und mit einer Formel verarbeitet. Die Formel lautet:

$$\$Score_Competition\$*0.3+\$Score_ProductValue\$*0.2+\$Score_SalesDevelopment\$*0.5$$

Was sagt uns das aus?

Die Formel berechnet einen Gesamtscore aus drei Faktoren, wobei Sales Development mit 50 % den größten Einfluss, gefolgt von Wettbewerb (30 %) und Produktwert (20 %) hat. Das bedeutet, dass die Verkaufsentwicklung strategisch am wichtigsten bewertet wird und den Gesamterfolg am stärksten bestimmt.

Danach wird bei dem Rule Engine mit:

$$\$Score_Overall\$ > 1 \text{ AND } \$Score_Overall\$ \leq 2 \Rightarrow 0.06$$

$$\$Score_Overall\$ > 0.5 \text{ AND } \$Score_Overall\$ \leq 1 \Rightarrow 0.03$$

$$\$Score_Overall\$ > -0.5 \text{ AND } \$Score_Overall\$ \leq 0.5 \Rightarrow 0$$

$$\$Score_Overall\$ > -1 \text{ AND } \$Score_Overall\$ \leq -0.5 \Rightarrow -0.03$$

$$\$Score_Overall\$ \geq -2 \text{ AND } \$Score_Overall\$ \leq -1 \Rightarrow -0.06$$

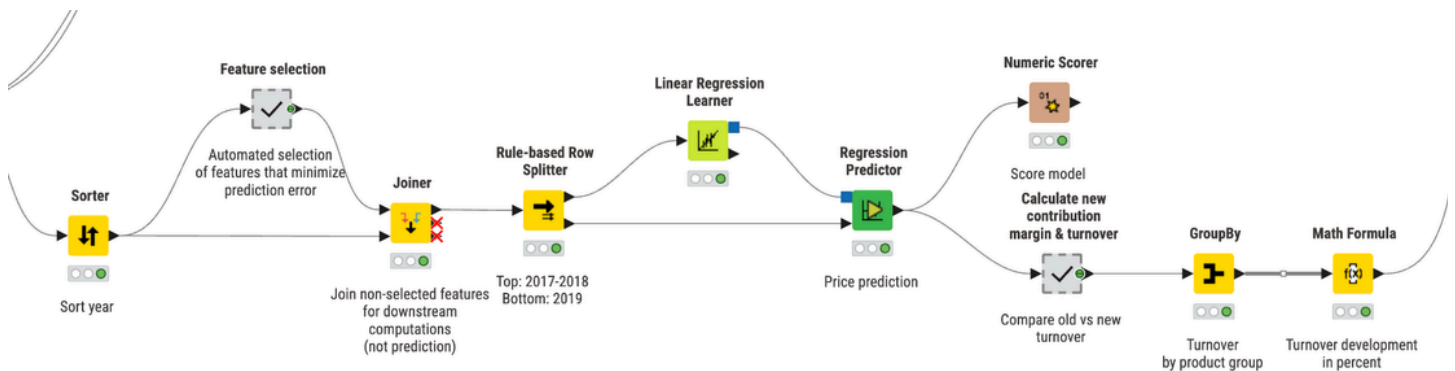
bestimmt. Das sagt aus, dass der Gesamtscore durch die Rule Engine in feste positive oder negative Anpassungswerte umgewandelt, die als konkreter Steuerungs- oder Entscheidungsfaktor dienen bestimmt wird.

Mithilfe dessen wurde der neue sales Preis kalkuliert und ausgewertet. Dies wurde wiederum einmal in eine Excel Datei geschrieben und einmal weitergeleitet zu einem Group By. Die Wirkung davon war einfach nur das die Tabelle nach eindeutigen Werten ausgewählter Spalten gruppiert wurde, wobei für jede einzigartige Kombination eine Zeile entsteht und die übrigen Spalten entsprechend definierter Aggregationsregeln zusammengefasst werden. hierbei handelte es sich um die Produktkategorie.

Nach dem wurde eine weitere Mathematische Formel verwendet und zwar:

$(\frac{\$Turnover_new}{\$Turnover_current} - 1) * 100$. - > dies beschreibt die prozentuelle Veränderung des Umsatzes.

Bevor es aber zusammengefügt wird, wurde auf der andern Seite die Preise prognostiziert, sowie als Lineare Regression dargestellt und analysiert.



Dafür musste es zuerst einen Sorter durchlaufen, welcher nach Jahren sortierte. Der Workflow zeigt anhand eines fiktiven E-Commerce-Datensatzes zwei Preisoptimierungsansätze wertbasierte Preisgestaltung und lineare Regressionsmodelle mit dem Ziel, durch Datenanalyse von Produkt-, Bestell- und Wettbewerbsinformationen den Umsatz zu steigern. Nach dem wurde es zusammengefügt und nach Jahren gesplittet. Oben waren 2017-2018 und unten 2019. Mithilfe dessen konnten wir einen Learner verwenden, dies Berechnet ein multiples lineares Regressionsmodell, indem eine Zielspalte als abhängige Variable und mehrere weitere Spalten als unabhängige Variablen ausgewählt werden.

Der Regression Predictor, wie schon im Namen steht profiziert die Regression des Models. Beim Numeric Scorer konnte dann eine Vorhersagung getroffen werden.

RowID	Prediction_Sales_Price
24	1
R^2	1
mean ε	0.248
mean ε	3.701
root mε	1.924
mean ε	0.228
mean ε	0.006

Gleichzeitig wurde eine Optimierung durch den Node Calculate new Contribution vorgenommen welcher zuerst noch durch eine Gruppierung (nach Produktkategorie) überlief, sowie erneut durch die selben mathematischen Formel für die prozentuelle Veränderung des Umsatzes, bevor es zum Endpunkt gelangte und

zwar zum Price Optimazition View. Dies ist ebenfalls ein Node was genau für solche Prognosen vorher gesehen sind.

Es hilft uns das alles zusammen zu führen und in einem Dokument zu veranschaulichen.

Zusammenfassung

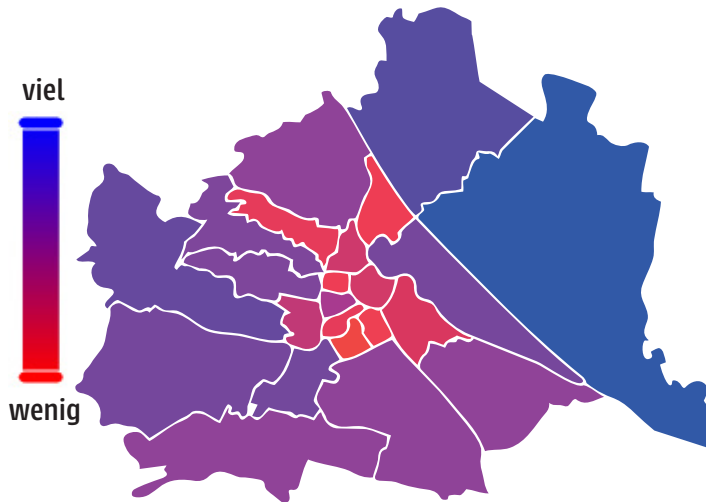
Die Übung war vor allem am Anfang besonders schwer da wir keine Vorerfahrung mit diesem Tool besitzen. Es hat auch während der durchführung leider vieles nur schwer oder garnicht funktioniert. Jedoch ging es nach einer Zeit und dann wurde es immer leichter die Vorteile davon zu sehen. In machen Fällen kommt es einem so vor als wäre es mit Excel viel einfacher, doch der schein trübt. Den genau für solche Sachen war es extremst praktisch das es bereits Nodes gibt die das alles vereinfachen, sachen miteinander verbinden werden können, prognosen getroffen werden können und das viel einfacher veranschaulicht wird als nehmen wir mal an in Excel. Man muss trotzdem zu dem Punkt kommen das es Sinn macht, dies ist nur bei wiederholten Projekten der Fall, wie Prognosen stellen, mit verschiedenen Zahlen Dokumente schreiben oder Machine Learning.

Wenn diese Punkte bedacht werden ist die Nutzung dieser Applikation sehr sinnvoll und sollte gelehrt werden. Kommt es jedoch nicht zu genügend nutzen, ist es sehr schwer in das ganze reinzukommen und ebenfalls funktioniert es nicht immer. Es traten sehr oft Fehler auf, dass zB keine Daten mehr hochgeladen werden durften.

Hunde in Wien

Andreas Juchelka 03.03.2026

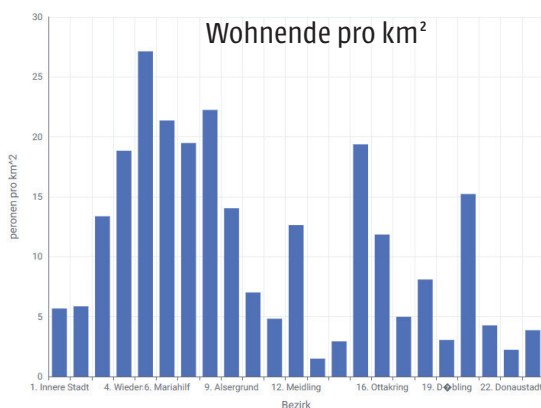
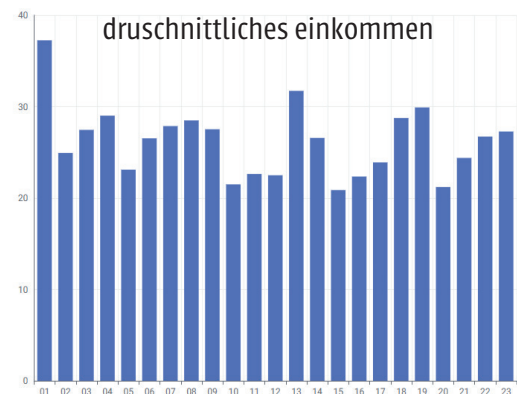
Hundesackerlspender pro Hund



In Wien gibt es insgesamt 33 793 Hunde. Die meisten gibt es im 22. Bezirk (2730) die wenigsten im 8. Bezirk (700). Die Anzahl der Hundesackerlspender ist relational fast gleich mit der Anzahl der Hunde wodurch der 22. Bezirk die meisten (545) und der 8. Bezirk die wenigsten hat (54). Grundsätzlich kann gesagt werden dass es in Wien ca. pro 8 Hunde einen Hundesackerlspender gibt. Doch es gibt natürlich auch Ausreißer bei dieser Rechnung wie zum Beispiel den 5. Wiener Gemeindebezirk indem das Verhältnis zwischen Hunde und Spendern ein Maximum von 16 zu 1 erreicht. Aus den Zahlen lässt sich auch herauslesen, dass Wien Donaustadt ein besonders Hundefreundlicher Bezirk ist in ihm wohnen nicht nur die meisten Hunde sondern auch das Verhältnis zwischen Hundesackerlspender und Hunden ist kleinste (1 Spender pro 5 Hunde).

Weiters kann man auch diese Daten mit dem durchschnittlichen Einkommen der Bezirke in Vergleich setzen. Haben reichere oder ärmere Menschen mehr Hunde?

Die Daten zeigen einen schwachen aber eindeutigen Trend an, dass reiche Bezirke eher weniger Hunde haben, arme Bezirke mehr. Jedoch ist bei diesen Daten zu beachten, dass diese Feststellung nicht als allgemein richtig wahrgenommen werden kann, denn solche Daten können leicht falsch interpretiert oder missverstanden werden, wenn man zum Beispiel bedenkt, dass wir bei dieser Annahme die Größe der Bezirke und die Einwohner pro Bezirk nicht berücksichtigt wurden.



In Bezirken mit hoher Bevölkerungsdichte, wie z.B. dem 5. Bezirk mit 27.133 Einwohnern/km², gibt es nur 1045 Hunde, während weniger dicht besiedelte Bezirke wie der 22. Bezirk mit nur 2,23 Einwohnern/km² 2730 Hunde haben. Überraschend ist, dass der Bezirk mit dem höchsten Einkommen (1. Bezirk, 37,25 Tsd.) nur 734 Hunde hat, obwohl er nicht die höchste Dichte aufweist, und dass einige Bezirke mit mittlerer Dichte, wie der 10. Bezirk mit 7.012 Einwohnern/km², trotzdem relativ viele Hunde (2222) haben. Insgesamt zeigt sich: Bevölkerungsdichte ist der stärkste Faktor für die Hundezahl, aber Ausnahmen verdeutlichen, dass auch andere lokale Gegebenheiten Einfluss haben.

Die Hundezahl hängt vor allem von der Bevölkerungsdichte ab: dichter besiedelte Bezirke haben weniger Hunde, dünn besiedelte deutlich mehr. Einkommen spielt nur eine geringe Rolle, Ausnahmen bestätigen aber, dass lokale Gegebenheiten zusätzlich Einfluss haben.

<https://www.istockphoto.com/de/vektor/zeiger-umriss-wien-im-blauen-umriss-gm1268240419-372218702>

<https://www.data.gv.at/datasets/71edef44-9d6c-4042-ab71-7207dc930ba7?locale=de>

<https://www.data.gv.at/datasets/98908792-f315-4834-9b5e-20fed66cbe5a?locale=de>

<https://www.data.gv.at/datasets/d76c0e8b-c599-4700-8a88-29d0d87e563d?locale=de>

<https://stp.wien.gv.at/viennaviz/anonymous/embed.html?id=e11e37c3-257a-4988-a919-3f04f24fa233&status=published>

KNIME vs Power BI

Vergleich der zwei Datenanalyse Möglichkeiten

Eine effektive Kombination für Datenanalyse und Visualisierung

KNIME

Beschreibung:

KNIME (Konstanz Information Miner) ist ein leistungsstarkes Open-Source-Tool für ETL (Extract, Transform, Load) und Data Science. Es ist besonders nützlich für komplexe Datenoperationen, die in traditionellen Tools schwer zu realisieren sind.

Komplexität:

KNIME kann mit sehr heterogenen Datenquellen umgehen und ist in der Lage, komplexe Berechnungen durchzuführen. Dies ist besonders vorteilhaft, wenn Daten aus verschiedenen Quellen integriert und angereichert werden müssen.

Flexibilität:

Dank seiner Open-Source-Natur und einer umfangreichen Bibliothek an Erweiterungen bietet KNIME eine hohe Flexibilität bei der Datenaufbereitung und -transformation.

Nachteil:

KNIME hat eine steile Lernkurve, da die Vielzahl an Funktionen und Erweiterungen für Anfänger überwältigend ist. Bei der Verarbeitung sehr großer Datenmengen können Leistungsprobleme auftreten, die oft spezielle Hardware erfordern. Zudem ist die Benutzeroberfläche nicht so intuitiv wie bei neueren Tools, und die Community-Unterstützung ist im Vergleich zu kommerziellen Lösungen weniger umfassend. Ebenfalls, ist es teilweise nicht möglich Erweiterungen zu installieren welche Essenziell sind für weitere Datenanalyse.

Einsatzgebiet:

KNIME ist ideal, wenn es darum geht, Daten zu bereinigen, zusammenzuführen oder sie durch statistische Modelle wie Machine Learning zu bereichern.

Power BI:

Beschreibung:

Power BI ist ein Business-Intelligence-Tool von Microsoft, das darauf ausgelegt ist, Daten in aussagekräftige Informationen zu übersetzen. Es ist besonders geeignet für die Erstellung von interaktiven Dashboards und Berichten.

Visualisierung:

Power BI bietet eine exzellente Auswahl an Diagrammen und Tools, um interaktive Dashboards zu erstellen, die einfach geteilt werden können.

Integration:

Wenn Ihr Unternehmen bereits im Microsoft-Ökosystem arbeitet, integriert sich Power BI nahtlos und bietet eine hervorragende Benutzererfahrung.

Einsatzgebiet:

Power BI glänzt, wenn es darum geht, Ergebnisse zu präsentieren, Trends schnell zu erfassen und Self-Service-Analysen für Anwender ohne tiefe IT-Kenntnisse bereitzustellen.

Zusammenfassung:

Wenn es mehr um Machine Learning und ähnliches geht, würde ich KNIME empfehlen, jedoch muss man sich dabei dessen bewusst sein das sehr viel Zeit nur schon in die Vorbereitungsphase und Verständniss mit dem Produkt, fließen wird. Power BI ist jedoch im Gegensatz mit seinen vielen Diagrammen und Tools für Reports oder Präsentationen viel verpreschend. Vor allem eignet es sich mehr wenn man schon einen Microsoft Zugang hat. Falls man nicht mehr Geld ausgeben mag und nach einer günstigen Open Source Möglichkeit sucht, ist KNIME die perfekte Wahl.