Nachvollziehbare Schritte
Projektarbeit

"Kundensegmentierung"

1. Kurze Darstellung des Problembereichs / Aufriss des Themas

1.1 Inhaltlich

Die Segmentierung von Kunden ist ein Schlüsselinstrument zur Optimierung von Marketingstrategien, Kundenmanagement und Steigerung des Unternehmensgewinns. Sie ermöglicht es Unternehmen, ihr Publikum besser zu verstehen und ihre Handlungen entsprechend den Bedürfnissen und Vorlieben anzupassen.

In dieser Projektarbeit wurden die Kunden einer Autofirma mit KNIME auf einige Zielgruppen geteilt.

Ziel dieser Arbeit

Alle Kunden in Gruppen einteilen, um die Kategorie der Hauptabnehmer zu verstehen und neue Marketingstrategien zu entwickeln.

Quelle

Customer Segmentation: https://www.kaggle.com/datasets/vetrirah/customer

1.2 Begründung des Themas

Die Segmentierung von Kunden ist ein Schlüsselelement einer erfolgreichen Marketingstrategie und Kundenmanagement für jedes Unternehmen. Sie ermöglicht es Unternehmen, ihre Zielgruppe besser zu verstehen, ihre Bedürfnisse und Vorlieben zu identifizieren und ihre Produkte, Dienstleistungen und Kommunikation entsprechend anzupassen, um die Kundenzufriedenheit zu maximieren. Dies führt zu effektiveren Marketingkampagnen, erhöhter Kundenloyalität und gesteigerter Unternehmensrentabilität. Somit spielt die Kundensegmentierung eine wichtige Rolle bei der Erreichung von Wettbewerbsvorteilen und dem erfolgreichen Wachstum eines Unternehmens.

2 Nachvollziehbare Schritte

2.1 Der Stand der Forschung / Auswertung der vorhandenen Literatur / Tutorials ...

Dieser Datensatz wurde bisher für mehrklassige Klassifikationsprobleme verwendet.

2.2 Methode

2.2.1 Vorbereitung

In dieser Projektarbeit geht es um die Kunden einer Autofirma. Das Unternehmen will alle Kunden einteilen, um neue Marketingstrategien zu entwickeln.

Die Originaldaten wurden in zwei Dateien Train.csv und Test.csv bewährt, die von der Quelle kopiert wurden. Die Daten wurden gereinigt. Es gibt die fehlenden Werte. Früher wurde der Datensatz für die mehrklassige Klassifikation verwendet. Die letzte Spalte in der Train.csv enthält Informationen über die Klassen.

2.2.2 Bibliothek importieren

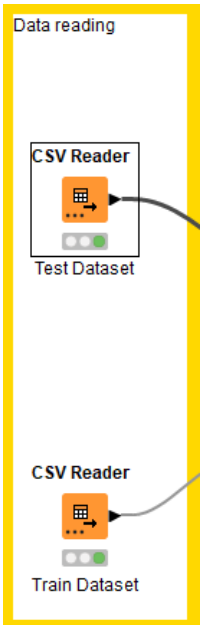
Nur Standardknoten werden benutzt.

2.2.3. Workflow.

Die Hauptschritte:

- Dateien laden,
- Preprocessing,
- Clustering,
- Datenspeicherung und Visualisierung.
- Statistiken für einen der Cluster.

2.2.4. Dateien laden.



Daten aus zwei csv-Dateien werden mit der Knote CSV Reader:

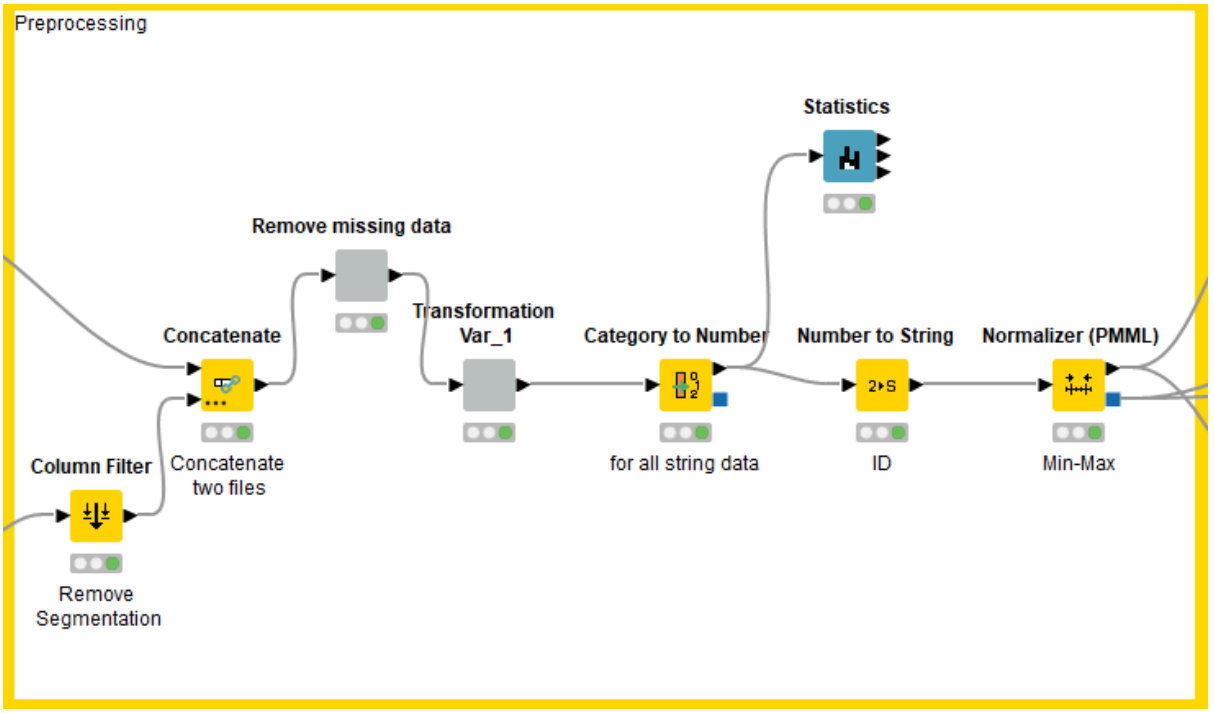
- Deteiweg geben,
- ein Komma als Trennzeichen und das Vorhandensein eines Headers in den Dateien angeben.

Die Variablen:

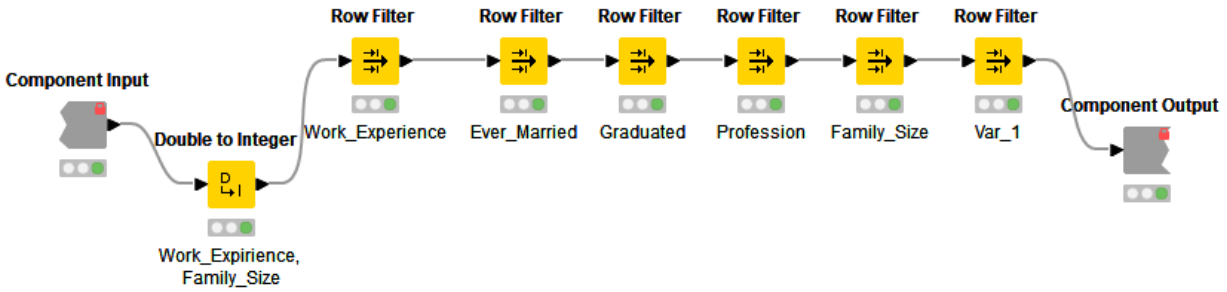
- ID -Kundennummer
- Gender - Geschlecht
- Ever_Married - verheiratet oder Single
- Age -Alter
- Graduated - Bidung (ja/nein)
- Profession -Beruf
- Work_Experience- Wie viele Jahre arbeitet der Kunde?
- Spending_Score -Ausgabebewertung (niedrig, mittel oder hoch)
- Family_Size - Wie viele Leute sind in der Familie?
- Var_1 - Kundenverhalten

Variablen Work_Experience und Family_Size in den Format Fließkommazahl.

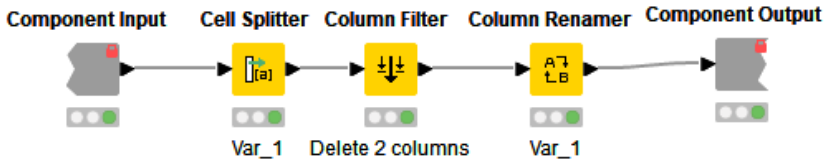
2.2.5. Preprocessing



- Variable Segmentation aus Test.csv Datei entfernen (Column Filter). Clustering ist eine Methode des unüberwachten Lernens.
- zwei Dateien zusammenführen (Concatenate).
  Insgesamt wurden 10695 Zeilen hochgeladen. Man kann die fehlenden Werte entfernen, weil Daten genug für das Clustering sind.
- Component 'Remove missing values'. Die Variable Work_Experience und Family_Size auf Ganze Zahlen wechseln (Double to Integer) und alle fehlenden Werte nacheinander entfernen (Row Filter). Nach dem Entfernen fehlender Werte sind noch 8819 Zeilen übrig.



- Component 'Transformation Var_1'. Die Werte in Var_1 sind Zeichenketten, die auf Ganzzahlen geändert werden müssen. Spalte teilen-> zwei Spalte mit Zeichenketten entfernen -> Spalte mit Ganzzahlen in Behavior umbenennen.

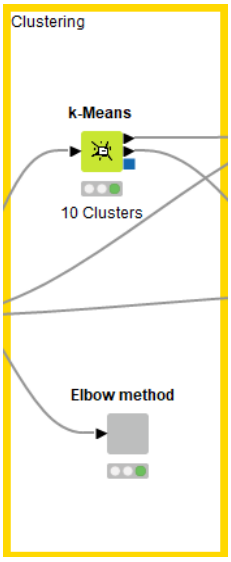


- Alle Daten in Zeichenketten-Format in Ganzzahl-Format konvertieren (Category to Number).
- Variable ID kann die Ergebnisse nicht beeinflussen, deshalb werden die Werte in Zeichenkette-Format konvertiert.
- alle Werte mit Min-Max skalieren (von 0 bis 1)(Normalizer(PMML))
- Die Knoten Statistik zeigt die gemeinsame Statistik der Daten.

Port Output | Port 0 | Load data | Rows: 10, Columns: 16

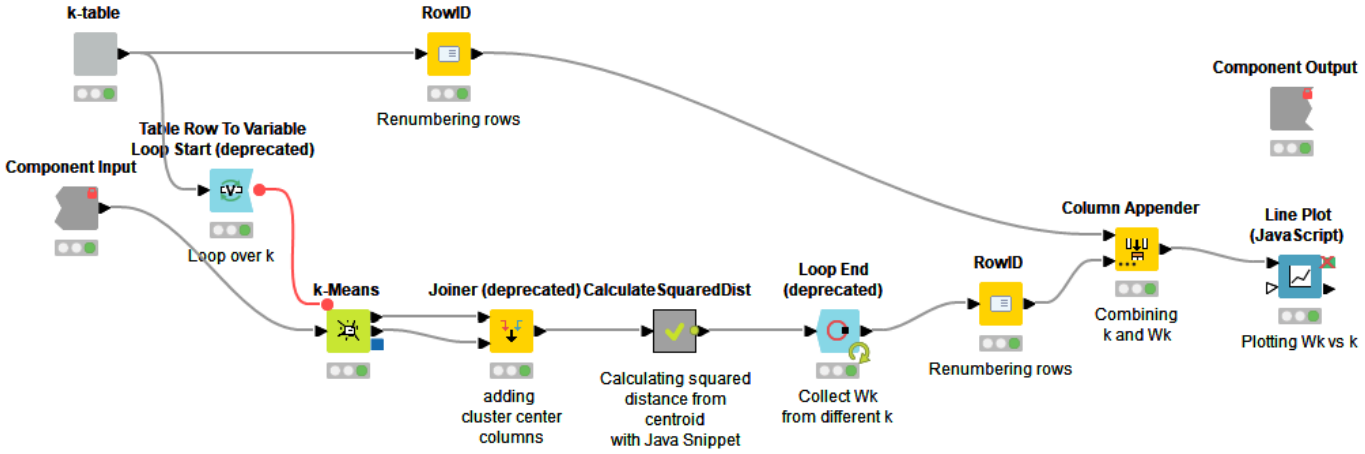
ID	Column	Min	Max	Mean
Gender	Gender	0	1	0.5511962807574532
Ever_Married	Ever_Married	0	1	0.4085497221907243
Age	Age	18	89	43.51785916770636
Graduated	Graduated	0	1	0.36568771969611163
Profession	Profession	0	8	3.4709150697357942
Work_Expe...	Work_Expe...	0	14	2.610159882072802
Spending_S...	Spending_...	0	2	0.5448463544619581
Family_Size	Family_Size	1	9	2.840117927202632
Behavior	Behavior	1	7	5.175530105454148

2.2.6.Clustering

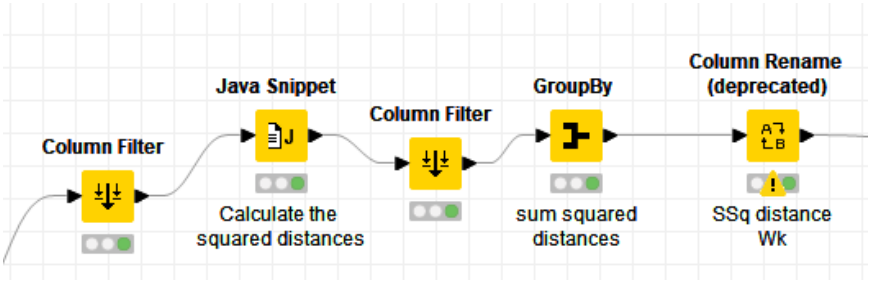


Der Algorithmus k-Means für Clustering ausgewählt wird.Für die Analyse muss vorab festgelegt werden, in wie viele Gruppen der Datensatz unterteilt wird.
Die optimale Anzahl von Clustern im K-Means-Algorithmus wird mithilfe der Ellbogen Methode bestimmt. Diese basiert auf der Analyse der Änderung der innerhalb der Cluster-Summe der Quadrate (Within-Cluster Sum of Squares, WCSS) bei Änderung der Anzahl der Cluster.

1. *Elbow method*

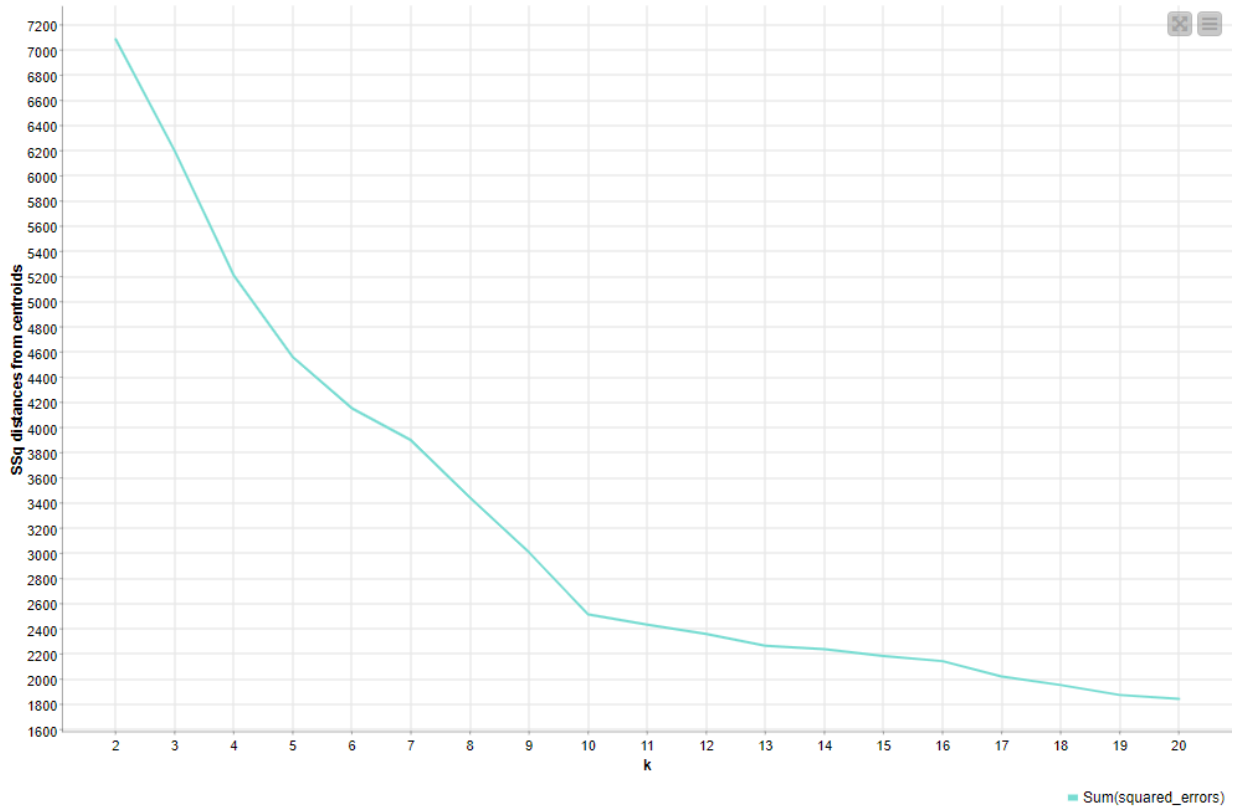


- in k-table die maximale Anzahl von Clustern eingeben
- die Daten auf die Clustern in der Schleife für Clustern von 2 bis 20 unterteilen.
- für jede Zeile ein Zentroid bestimmen und seine Daten hinzufügen
- die Summe der quadratischen Abstände zwischen dem Objekt und der Zentrode mit Java Snippet für jedes Modell aus der Schleife rechnen.



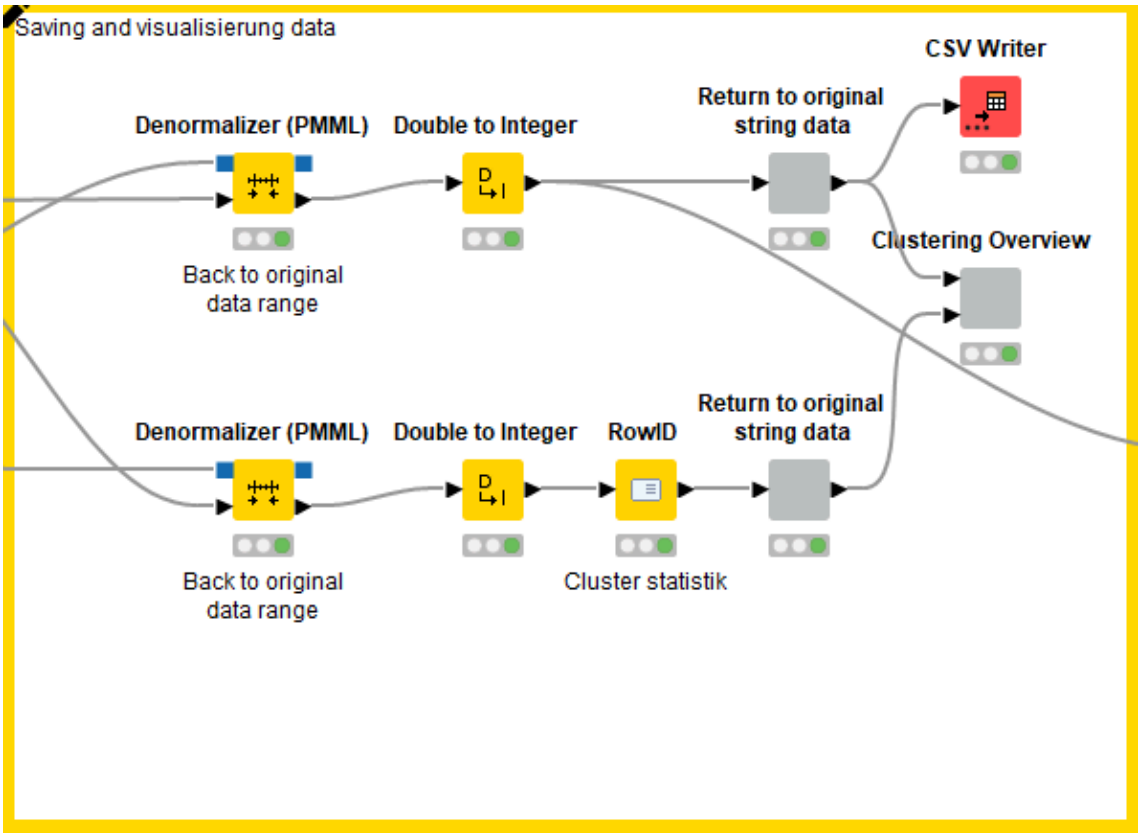
- ‘’Column Appender’’ verbindet die Ergebnisse der Schleife und die entsprechende Clusteranzahl.
- Die Abhängigkeit der Distanzsumme von der Clusteranzahl wird grafisch dargestellt.

Die Steigung der Kurve ändert sich bei 10 Clustern stark. Objekte befinden sich ziemlich nahe am Zentroid. Eine weitere Erhöhung der Clusteranzahl verändert den Abstand geringfügig.

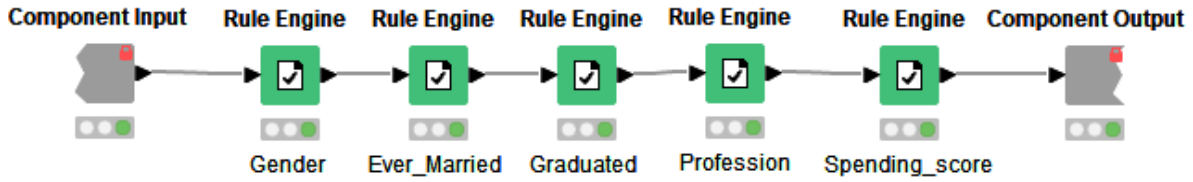


2. ***Clustering k-means***
   Kundendaten sind in 10 Gruppen mit k-means unterteilt.

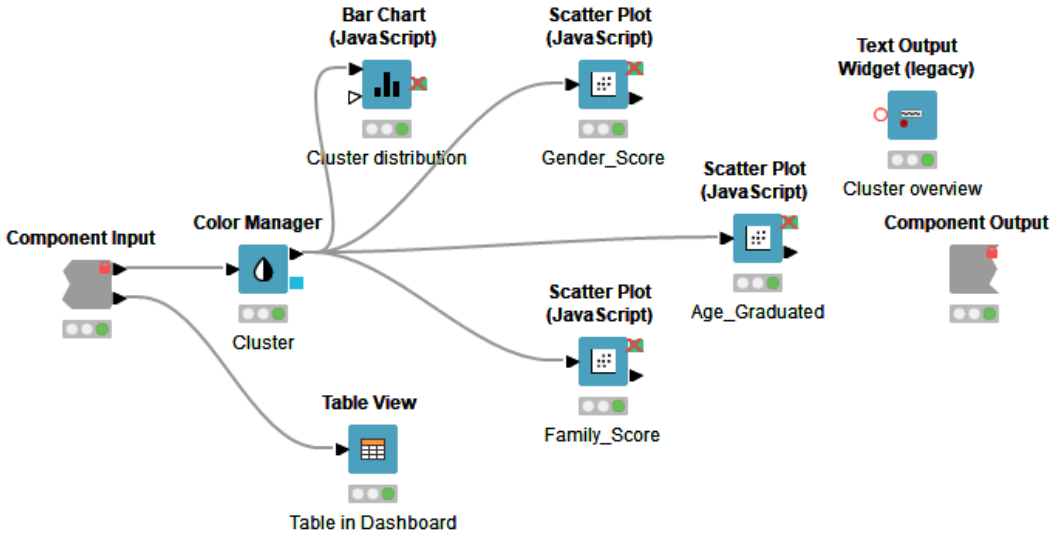
2.2.7. Datenspeicherung und Visualisierung.



- die Daten auf den ursprünglichen Maßstab zurückbringen (Denormalizer) und Fließkommazahlen ins Ganze Zahl-Format konvertieren (Double to Integer).
- Die Knote Denormalizer (PMML), die oben ist, bringt alle Zeilen mit den entsprechenden Cluster zurück. Die Knote Denormalizer (PMML), die unten ist, bringt die Informationen über die Clastern von 0 bis 9.
- Die Variablen Gender, Ever_Married, Graduated, Profession und Spending_Score werden in Zeichenketten Kategorien zurückzukehren (‘Component ‘’Return to original string data’’):

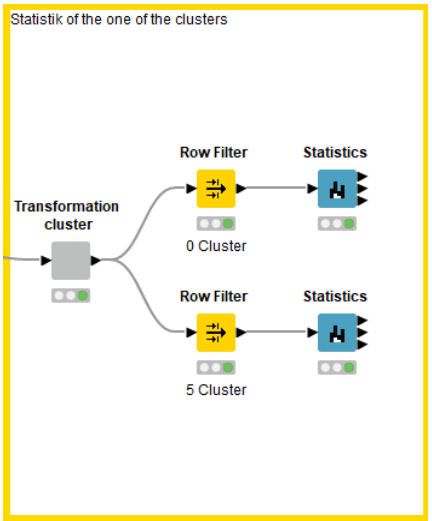


- die Ergebnisse in csv-Datei speichern (CSV Writer).
- Daten in dem ursprünglichen Format werden an das Dashboard übertragen (Component Clustering Overview):



Für die Visualisierung der Tabelle wird der Knoten ‘Table View’ verwendet.

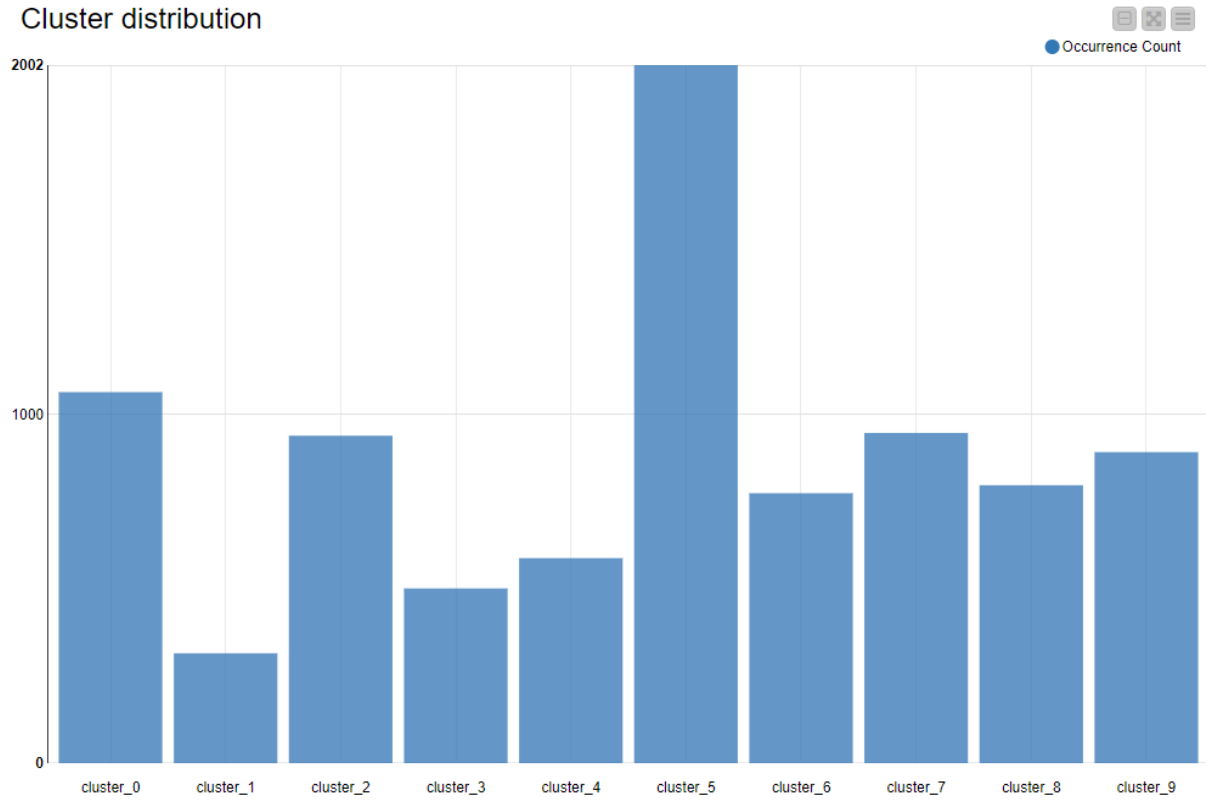
2.2.8. Statistiken für einen der Cluster.



-Die Zeilen für den entsprechenden Cluster auswählen (Row Filter).
- Statistiken anzeigen (Statistics)
Daten sind im Ganzen Zahl-Format.

2.3 Ergebnisse

Clusters:
0 - unverheiratete, gebildete Frauen mittleren Alters mit geringem Kaufverhalten.
1 - verheiratete, gebildete Männer mittleren Alters mit durchschnittlichen Ausgaben und ohne Kinder.
2 - verheiratete, ungebildete Männer mittleren Alters mit mittleren Ausgaben und mit Kindern.
3 - verheiratete, gebildete, ältere Männer mit hohen Ausgaben.
4-verheiratete, gebildete Männer mittleren Alters mit geringem Kaufverhalten und ohne Kinder.
5 - verheiratete, gebildete Frauen mittleren Alters mit mittleren Ausgaben und mit Kindern.
6 - verheiratete, gebildete Männer mittleren Alters mit mittleren Ausgaben und mit Kindern.
7 - unverheiratete, ungebildete Männer mit geringem Ausgabenniveau (Studenten oder Rentner).
8 - unverheiratete, ungebildete Männer mittleren Alters mit geringem Ausgabenniveau.
9 - unverheiratete, ungebildete Frauen mit geringem Ausgabenniveau (Studentin oder Renten).



Die größte Kundengruppe sind die verheirateten, gebildeten Frauen mittleren Alters mit mittleren Ausgaben und mit Kindern. Es gibt 2-mal mehr von ihnen als jede andere Gruppe (Cluster 5).
Die Kategorie 5 könnte Familienautos mit erweiterten Sicherheitsfunktionen und komfortablen Innenräumen, günstigen Finanzierungsmöglichkeiten und Rabatte für Familienkäufer angeboten werden.

Die unverheirateten, gebildeten Frauen mittleren Alters mit geringem Kaufverhalten liegen an zweiter Stelle (Cluster 0).
Dieser Gruppe könnten wirtschaftliche und preisgünstige Automodelle, flexible Kreditbedingungen, ein Programm zum Austausch von alten Autos, Garantieleistungen und technische Beratung angeboten werden.

2.4 Ausblick

In dieser Projektarbeit wurden die Kunden einer Autofirma auf 10 Zielgruppen geteilt. Die größten Kundengruppen sind die verheirateten Frauen mit mittleren Ausgaben und die unverheirateten Frauen mit geringem Kaufverhalten. Diesen Gruppen werden Produkte und Marketingstrategien angeboten, die für sie von Interesse sein könnten.