

Hausaufgabe 2

Contents

1	Simulationsspiel	2
2	Aufgaben	2
2.1	Aufgabe 1 - Datenanalyse (10 Punkte)	2
3	Formale Anforderungen	3

1 Simulationsspiel

Die nachfolgenden Aufgaben beziehen sich auf das Simulationsspiel. Es gelten weiterhin die Informationen, die bereits in Hausaufgabe 1 vorgestellt wurden. Es fand keine Simulation von neuen Monaten statt. Stellen Sie sich stattdessen vor, dass Sie nach den letzten Analyseergebnissen (HA 1) nun weitere Analysen durchführen möchten, die Ihre Entscheidungsfindungen weiter verbessern sollen.

2 Aufgaben

2.1 Aufgabe 1 - Datenanalyse (10 Punkte)

Sie haben nun etwas Erfahrung auf dem Aufgabengebiet der Distribution von "Limonalytics" und sollen fuer die naechsten Monate die Beauftragung von Logistikdienstleistern planen. Dazu ist es notwendig, die Nachfrage vorherzusagen.

Hinweis: Die Hinweise aus HA 1 zur allgemeinen Bearbeitung der Aufgaben gelten weiterhin.

Daten vorbereiten

- 1) Aggregieren Sie die Verkaufszahlen so, dass Sie eine Tabelle mit der Nachfrage je Monat **je Region** bekommen (fuer alle Produkte zusammen einschliesslich Lost Sales! Sie brauchen diese Daten, um die Nachfrage in einer Region abzuschätzen). Bereinigen Sie die Daten anschliessend von unnuetzen Daten. Die Daten sollten in einer neuen Variable mit dem Namen "Demand" gespeichert werden. Die Spalte, welche im DF transactions noch "Sales" heisst, soll im neuen DF Demand "Demand" heissen. Geben Sie nur den Tabellenkopf aus. Bewertungsrelevant: Code, Output.

Hinweis: Fuer die folgenden Aufgaben bietet es sich an, eine Variable "Periode" zu erzeugen, die aus dem Jahr und dem Monat (in dieser Reihenfolge und jeweils durch eine Zahl repraesentiert) besteht. Siehe dazu die Dokumentation unter diesem Link und die Dokumentation unter diesem Link.

- 2) Wandeln Sie die aggregierten Demand-Daten vom Long-Format in das Wide-Format um. (Fuer den vorliegenden Fall gibt es im Long-Format drei Spalten: Periode, Region und Demand. Im Wide-Format gibt es im vorliegenden Fall sechs Spalten: Periode, Demand in Japan, Demand in Peking, Demand in Phlppn, Demand in Shangh, und Demand in Skorea.) Nutzen Sie dazu die Reshape()-Funktion. Infos unter diesem Link (<- bitte anklicken). Geben Sie nur den Tabellenkopf aus. Bewertungsrelevant: Code, Output.
- 3) Wandeln Sie die aggregierten Verkaufszahlen in den Datentyp time-series mit Frequenz = 12 um. Jede Demand-Spalte (jeder Vektor) soll dabei in eine Time Series umgewandelt werden. Dafuer benoetigt es nur 5 Befehle. Fuer diese Aufgabe brauchen Sie nichts ausgeben. Bewertungsrelevant: Code.

Modellierung vorbereiten

- 4) Visualisieren Sie den Nachfrageverlauf aller Regionen in einem einzigen ggplot. Achten Sie auf uebersichtliche Achsenbeschriftungen und Legenden. Heben Sie exemplarisch die Region Peking hervor, damit man den Nachfrageverlauf dort gut erkennt (die anderen Regionen muessen nicht unbedingt gut erkennbar, jedoch theoretisch unterscheidbar sein sein). Bewertungsrelevant: Output.

Hinweis: Wenn Sie ggplot verwenden, kann es notwendig sein, dass Sie bei den aesthetics zusaetzlich `group = 1` angeben muessen, damit Sie einen Plot sehen. Eventuelle Fehlermeldungen sind nicht bewertungsrelevant.

- 5) Betrachten Sie die Visualisierung und nehmen Sie zu moeglichen Trends und Saisonalitaeten fuer die exemplarische Region Stellung. Erachten Sie eine Zeitreihenanalyse in diese Fall als sinnvolle Methode, um den Nachfrageverlauf zu modellieren? Angenommen, dies ist der Fall: Welche Annahme muessen sie dann treffen, um die Ergebnisse der Modellierung fuer Ihr Produkt nutzen zu koennen? Bewertungsrelevant: Kommentar.

Modellierung

- 6) Nutzen Sie nun Ihre exemplarische Region weiter. Fuehren Sie mit Hilfe der ets()-Funktion eine Zeitreihenanalyse durch. Erstellen Sie ein Modell, indem Sie die automatische Festlegung der Modellparameter nutzen (model

= "ZZZ"). Geben Sie (1) eine Zusammenfassung des Modells, (2) die urspruengliche Zeitreihe und (3) die Werte der Residuen aus. Bewertungsrelevant: Code, Output.

- 7) Berechnen Sie die durchschnittliche monatliche Abweichung der Modellwerte von den Originalwerten im Jahr 2020. Interpretieren Sie das Ergebnis indem Sie von dem hypothetischen Fall ausgehen, dass im Jahr 2020 das Modell bereits in einem beispielhaften Unternehmensprozess genutzt wurde. Bewertungsrelevant: Output, Kommentar.
- 8) Erstellen Sie eine Nachfragevorhersage fuer ein weiteres Jahr. Visualisieren Sie den Nachfrageverlauf sowie den Verlauf Ihres Prognosemodells in einem einzigen ggplot. Begründen Sie die Wahl der Visualisierung. Bewertungsrelevant: Code, Output, Begründung.
- 9) Bewerten Sie Ihr Modell aus Aufgabe 6 mit Hilfe von 4 verschiedenen Fehler-Kennzahlen, die Sie aus der Uebung kennen. Welche der Fehler-Kennzahlen halten Sie fuer geeignet, um die Guete des Modells zu bewerten? Bewertungsrelevant: Output, Kommentar.
- 10) Vergleichen Sie Ihr Modell insgesamt mit Ihren Vermutungen aus Aufgabe 5. Was stellen Sie fest? Bewertungsrelevant: Kommentar.
- 11) Ihr Kollege Gaspar und Ihre Kollegin Belinda diskutieren. Gaspar ist der Meinung, er haette ein besseres Modell entwickelt als Sie. "Ich habe die Modellparameter `model = "ANA"` genutzt und denke das ist die bessere Variante.", sagt er selbstueberzeugt in der Mittagspause. "Unsinn!", wirft Belinda beherzt ein. "Die Modellparameter `model = "ZZZ"` ergeben immer das optimale Ergebnis. Das ZZZ-Modell ist besser." Vergleichen Sie die Modelle. Wer hat Recht? Bewertungsrelevant: Output, Kommentar.

Hinweis: Gehen Sie unabhängig der Ergebnisse der letzten Aufgabe davon aus, dass das ZZZ-Modell das beste Modell ist und weiterhin genutzt wird.
- 12) Erstellen Sie fuer die uebrigen Regionen ebenso Modelle zur Nachfragevorhersage (d.h. 4 weitere Modelle). Nutzen Sie erneut die automatische Festlegung der Modellparameter (`model = "ZZZ"`). Berechnen Sie zudem fuer jedes Modell den MAE, um den *Vergleich der Modelle* zu ermoeglichen. Nehmen Sie dazu Stellung, welches Modell laut der bewertenden Information das "beste" der vier Modelle sei. Bewertungsrelevant: Output, Kommentar.

Abschluss

Hinweis: Fuer die Bearbeitung der folgenden Aufgaben muessen Sie keine neuen Modelle erstellen.

- 13) Ihre Chefin ist begeistert von Ihren 5 Vorhersagemodellen. "Das bedeutet ja, dass wir voraussagen koennen, wie viele Flaschen Limonade wir insgesamt fuer die zweite Haelfte von 2021 benoetigen!", sagt sie freudestrahlend. "Koennen Sie mir dafuer eine ungefaehre Zahl nennen?". Helfen Sie Ihrer Chefin. Bewertungsrelevant: Output, Kommentar.
- 14) Ihr guter Freund Olaf wohnt in Japan und besitzt dort einen der fuenf Supermaerkte, die Ihre Limonaden-Marke fuehren. Er wuerde gerne wissen wie viele Flaschen Limonade im Dezember durchschnittlich in seinem Supermarkt nachgefragt werden. Ausserdem wuerde er gerne wissen, ob er fuer den Dezember diesen Jahres mit mehr oder weniger Limonade rechnen soll. Helfen Sie Olaf. Bewertungsrelevant: Output, Kommentar.

3 Formale Anforderungen

Abzugeben ist ein R-Markdown Notebook als Rmd- sowie als PDF-Datei, welches Ihre Antworten auf saemtliche Aufgabenstellungen enthaelt und vollstaendig ausfuehrbar ist.

Fuer den Umgang mit den R-Markdown Notebooks beachten Sie bitte das Cheat Sheet unter diesem Link und das Reference Handbook unter diesem Link. Wie in der Vorlesung angesprochen, koennen Sie erfolglose Code-Fragmente im R-Notebook lassen, beachten Sie jedoch, dass diese auf `eval = FALSE` gesetzt sind.

Benennen Sie die das Markdown-Notebook und die PDF-Datei wie folgt:

SCA_SS2021_Gruppe###_HA2

Bitte versuchen Sie Ihre Texte sinnvoll zu strukturieren!

Spätester Abgabepunkt: 26.05.2021 / 10:00 Uhr

Viel Erfolg!