

FALLSTUDIE

Aufgabenstellung zum Kurs: DLBDBSC01 – Statistical Computing

INHALTSVERZEICHNIS

1. Aufgabenstellung.....	2
1.1. Aufgabenstellung 1: Neues Land für smarten Tutor.....	3
1.2. Aufgabenstellung 2: Der ideale Wohnort	4
1.3. Aufgabenstellung 3: Hilfe an der richtigen Stelle	5
1.4. Weiterführende Literaturhinweise und Quellen	6
2. Zusatzinformationen zur Bewertung der Fallstudie	6
3. Betreuungsprozess	7

1. AUFGABENSTELLUNG

Komplementär zum Skript des Kurses wird in der Fallstudie ein Verständnis entwickelt, wie Daten aus verschiedenen öffentlichen Quellen importiert und vorbereitet werden, damit eine statistische Analyse gemacht werden kann und daraus Schlussfolgerungen gezogen werden. Die Fallstudie ist eine der besten Strategien, um die gesamten Kompetenzen der *Data Science Venn Diagramm* (D Conway, 2013) zu üben.

Wähle für Deine Bearbeitung einer der folgenden Fallstudien aus. Bitte berücksichtige bei Deiner Bearbeitung die in der jeweiligen Fallstudie selbst beschriebene Aufgabenstellung.

Hinweis zum Urheberrecht und zur Plagiatsprüfung:

Es wird darauf hingewiesen, dass der IU Internationale Hochschule GmbH das Urheberrecht der Prüfungsaufgaben/Aufgabenstellungen obliegt. Einer Veröffentlichung der Aufgabenstellungen auf Drittplattformen wird ausdrücklich widersprochen. Im Falle einer Zuwiderhandlung stehen der Hochschule u.a. Unterlassungsansprüche zu. Zudem weisen wir darauf hin, dass jede eingereichte schriftliche Ausarbeitung mittels einer Plagiatssoftware überprüft wird. Wir empfehlen daher auch, keinesfalls ausgearbeitete Lösungen zu teilen, da dies den Verdacht eines Plagiates begründen kann.

1.1. Aufgabenstellung 1: Neues Land für smarten Tutor

Hintergrund: Die Firma *Little Startup* ist in Deutschland sehr erfolgreich mit ihrem *Geometry-Companion*, eine mobile-App, die Schüler:innen dabei hilft, mathematische Sachverhältnisse tiefer und lehrerkonform zu vertiefen. Die Firma denkt nun darüber nach, die App auch in anderen Ländern anzubieten. Dafür will sie wissen, welches Land für ein solche App geeignet wäre. Da die App vieles aus dem Internet bezieht, möchten die Kolleg:innen der Marketingabteilung wissen, welches Land dafür bereit wäre. Die Einführung der App in einem anderen Land setzt sprachliche, curriculare, sowie betriebliche Anpassungen voraus. Dafür ist es wichtig, dass auch in Deutschland normale Kriterien wie die Internetbandbreite in diesem Land geprüft werden.

Es ist Deine Aufgabe herauszufinden, in welchem Land die höchste Zahl der Haushalte dazu in der Lage wären, eine App zum Lernen zu nutzen.

Um die Menge von Haushalten zu ermitteln, die in den vergangenen Jahren einen Breitbandzugang nutzten, kann der Datensatz *Internet Access* der OECD verwendet werden (siehe OECD Internet 2020), sowie die Tabelle der Haushaltsausgaben (siehe OECD Household spending 2020) und eventuell andere Informationen (etwa die Daten aus Wikipedia oder WorldBank zur Bevölkerung oder Lokalisierung).

Aufgabenstellung:

1. Passe die erhaltenen Daten an, damit sie als Zeitreihe für jedes Land in R passen. Normalisiere die Haushaltsausgaben pro Einwohner:in. Präsentiere diese Zeitreihen in Grafiken.
2. Korrelieren Breitbandzugang und Haushaltsausgaben? (Durch die Zeit? Zwischen den Ländern?)
3. Benutze lineare Regression basierend auf den letzten Jahren, um die Indikatoren für die kommenden zwei Jahre zu berechnen. Zeige diese auf den Grafiken (erstellt unter 1.).
4. Basierend auf dem Vergleich der Länder kannst Du einen Preis vorschlagen (mit dem aktuellen Preis in Deutschland von 12€ im Jahr). Schlage eine Berechnungsmethode vor, die Du komplett mathematisch darstellst.
5. Mit diesen analysierten Daten formulierst Du nun die Empfehlung in Form einer Liste von zehn Ländern, deren Markt bereit sein könnte für ein Angebot des *Geometry-Companion*. Diese statistisch begründete Empfehlung erklärst Du Deinen Kolleg:innen. Die Begründung besteht aus Beobachtungen der Daten und abgeleitete Zahlen, die sich für die Entscheidung relevant sind; sie muss von Personen, die wenig statistische oder informatische Kenntnisse haben, verständlich sein.
6. Dazu sind alle technische Stufen, die Du gemacht hast zu dokumentieren, damit andere Kolleg:innen (und der/die Korrektor:in dieser Fallstudie) die Berechnungen auch durchführen können, die Analyse verstehen, und auch leicht anpassen könnten.

1.2. Aufgabenstellung 2: Der ideale Wohnort

Hintergrund: Du und Dein Lebenspartner/Deine Lebenspartnerin habt die Möglichkeit, Eure Berufe vollständig im Homeoffice auszuüben. Aus diesem Grund seid Ihr nun in der Wahl eures Wohnorts frei und möchtet umziehen. Wichtig ist Euch bei der Wahl Eures zukünftigen Wohnorts nur, dass er eine geringe Wettervariabilität aufweist.

Du suchst dafür Orte in der Nähe einer der Stationen vom Deutschen Wetterdienst, um die dortigen Klimadaten der letzten Jahrzehnte zu analysieren. Die Kriterien, die die geringe Wettervariabilität ("so wenig wie möglich soll sich ändern") definieren, bleiben Dir überlassen; diese Kriterien müssen mindestens drei Wetterdatenparameter einbeziehen. Um Ortschaften zu finden, an denen es wenig Variabilität gibt, wendest Du Heuristiken an, wie zum Beispiel: In den Bergen gibt es immer wieder viel Sonne oder das Meer reguliert die Temperatur am besten. Diese kannst Du dann in Deiner Ortssuche beim Wetterdienst verwenden und Orte damit suchen. Das Ergebnis dieser ersten Suche soll eine sortierte Liste von zehn möglichen Ortschaften sein, in der Daten von jedem Tag der letzten drei Jahre grafisch und numerisch dargestellt sind. Ermittle auch eine Vorhersage für alle Tage der nächsten drei Jahre anhand linearer Regression und füge diese hinzu.

Aufgabenstellung:

1. Besuche das zentrale Datenangebot des Deutschen Wetterdienstes, das CDC Portal (Deutscher Wetterdienst 2020), und erkundige Dich über die möglichen Dimensionen des Wetters an einem Tag. Entscheide für Dich, welche drei Dimensionen entscheidend sein sollen für die Auswahl des neuen Wohnorts (Beispiele: Feuchtigkeit, Windgeschwindigkeit oder Tageshöchsttemperatur) und erfinde eine Begründung für die Bedeutung dieser Dimension für Euer Wohlbefinden. Dies sind nun Eure Kriterien für Wettervariabilität.
2. Etabliere und schreibe die Heuristik der Wahl des Ortes nieder und suche anschließend die Kandidaten in der CDC Applikation.
3. Extrahiere alle nötigen Daten für diese Orte seit zehn Jahren. Beschreibe genau Dein Vorgehen, damit andere dieses nachvollziehen können.
4. Skizziere die Sortierung nach Variabilität, indem Du ein Verfahren beschreibst, wie Du zwei Orte in den Daten, die Du gesammelt hast, vergleichen kannst. Am einfachsten kannst Du das machen, wenn Du einen Variabilitätsscore als Funktion der früheren Daten berechnest.
5. Aus den gesammelten Daten ermittelst Du jetzt ein Analyseergebnis. Dafür schreibst Du ein R Skript und wendest es auf die Daten an. Präsentiere Dein Analyseergebnis in geeigneter Weise, indem Du ein weiteres R Skript schreibst. Füge hier auch Deine Wettervorhersage für die nächsten drei Jahre hinzu.
6. Dokumentiere alle technischen Schritte Deiner Berechnungen, damit andere Kolleg:innen (und der/die Korrektor:in dieser Fallstudie) die Berechnungen nachvollziehen und reproduzieren können und die Analyse verstehen oder auch leicht anpassen könnten.

1.3. Aufgabenstellung 3: Hilfe an der richtigen Stelle

Hintergrund: Die neu gegründete wohltätige Organisation *WealthyGrowth* finanziert Kampagnen für Wirtschaftswachstum in Entwicklungsländern. Diese Kampagnen beinhalten oft sowohl politische Arbeit als auch Werbekampagnen in der Bevölkerung. Ziel dieser Kampagnen ist die Steigerung des GDPs per capita. In einem ersten Brainstorming Deiner Arbeitsgruppe wurden folgende Ansätze/Hypothesen aufgestellt:

- a. Wichtigster Ansatzpunkt ist eine höhere Bildung der breiten Bevölkerung. Die Kampagne sollte eine Schulpflicht für weiterführende Bildung als Ziel haben.
- b. Zentral ist die Rolle der Frau. Je mehr Frauen erwerbstätig sind, desto größer ist die verfügbare gesamte Arbeitskraft eines Landes. Die Kampagne sollte dementsprechend die Stellung der Frau im Beruf stärken und für eine Erwerbstätigkeit von Frauen werben.
- c. Der wichtigste Faktor für die verfügbare gesamte Arbeitskraft eines Landes ist die Lebenserwartung der Bevölkerung. Die Kampagne sollte also Maßnahmen zur Erhöhung der Lebenserwartung fokussieren.

Eine erste Datenanalyse auf frei verfügbaren Daten soll schnellen Aufschluss über das Potential dieser Ansätze/Hypothesen geben.

Aufgabenstellung:

1. Lade die relevanten Indikatoren von der World Bank Open Data Webseite herunter. Darunter sollten sich die folgenden Datensätze für die Jahre 2000-2019 befinden: "GDP per capita", die "Fertility Rate" (als Indikator für den Anteil erwerbstätiger Frauen) und die "Life expectancy at birth". Weitere Indikatoren können hilfreich sein, um die Fragestellung zu beantworten.
2. Bereinige die Datensätze um Länder, für die in mindestens einem Indikator für zu wenige Jahre Daten zur Verfügung stehen (z.B. für weniger als 30% der Jahre).
3. Nutze lineare Regression, um die fehlenden Werte zu interpolieren.
4. Formuliere einen Ansatz, wie man für Kampagne c. die Lebenserwartung steigern könnte (wie in der Fallbeschreibung bereits für a. und b. getan). Schlage Indikatoren vor, mit denen man den Einfluss dieses Faktors auf die Lebenserwartung plausibilisieren kann.
5. Nutze Korrelationsanalysen, um die oben aufgestellten Hypothesen a.-c. zu diskutieren. Formuliere aufbauend auf den bisherigen Analysen eine Empfehlung für eine der genannten Kampagnen und/oder erläutere weitere Analysen, die vor einer Entscheidungsfindung notwendig wären.
6. Dokumentiere alle technischen Schritte Deiner Berechnungen, damit andere Kolleg:innen (und der/die Korrektor:in dieser Fallstudie) die Berechnungen nachvollziehen und reproduzieren können und die Analyse verstehen oder auch leicht anpassen könnten.

1.4. Weiterführende Literaturhinweise und Quellen

Cengage Learning: *Case Studies*.

<http://college.cengage.com/business/resources/casestudies/students/overview.htm>

Conway, D. (2013). *The Data Science Venn Diagram*.

<http://drewconway.com/zia/2013/3/26/the-data-science-venn-diagram>

Das Fallstudien-Handbuch der Harvard Business School Press (2008). *Business-Cases entwickeln und erfolgreich auswerten*.

Ellet, W. (2007). *The Case Study Handbook: How to Read, Discuss, and Write Persuasively About Cases*. Harvard Business School Press.

Ellet, W. (2018). *The Case Study Handbook, Revised Edition: A Student's Guide*. Harvard Business Review Press.

OECD (2020). *Internet access* (indicator). <https://data.oecd.org/ict/internet-access.htm>

OECD (2020). *Household spending* (indicator). <https://data.oecd.org/hha/household-spending.htm>

Pearson Education: *How to Analyze a Case Study*.

http://wps.prenhall.com/bp_laudon_essmis_6/21/5555/1422312.cw/content/index.html

2. ZUSATZINFORMATIONEN ZUR BEWERTUNG DER FALLSTUDIE

Bei der Analyse der Fallstudie und ihrer Bearbeitung müssen die im Prüfungsleitfaden aufgeführten Bewertungskriterien und Erläuterungen berücksichtigt werden.

Erfassung: Wie hast Du den Auftrag für Dich formuliert und die Ziele für Dich gesetzt. Bringen Deine Ziele die erwartete Entscheidung? Die Hinweise zu den Quelldatensätzen sind nicht sehr genau: Wie hast Du diese Hinweise zu konkreten Aktionen umgewandelt?

Vorgehen: Du hast Dich für einen Vorgang entschieden. Durch das Lesen des Skriptes bist Du in der Lage gewesen, Werkzeuge anzuwenden, die die Daten auswerten konnten. Wie gut ist die Anwendung? Könnte sie vereinfacht werden, ohne ihre Richtigkeit zu verlieren? Inwiefern sind die Werkzeuge und angewandte Konzepte *Best-Practice*?

Qualität: Die Erklärung, die Du geliefert hast und die Datendarstellung der Daten, die Du präsentiert hast, werden bezüglich ihrer Qualität evaluiert: Sind die Entscheidungsvorschläge nachvollziehbar von Kolleg:innen, die wenig Ahnung von Statistik oder gar die Quelldatensätze haben? Inwiefern ist die Präsentation sprachlich, fachlich, und statistisch auf einem hohen Niveau?

Ergebnis: In dieser Fallstudie haben wir ein Ergebnis erbracht: Die Empfehlung einer Entscheidung. Während andere Kriterien den ganzen Weg zur Empfehlung evaluieren, fokussiert sich dieses Kriterium auf der Empfehlung (Anwendbarkeit und Vollständigkeit).

Formalia: Einhaltung der formalen Vorgaben

Richtigkeit: Korrektheit der Rechtschreibung und Interpunktion

Sprache: Qualität des sprachlichen Ausdrucks und Angemessenheit des Sprachstils für eine wissenschaftliche Arbeit

3. BETREUUNGSPROZESS

Für die Betreuung der Fallstudie stehen grundsätzlich mehrere Kanäle offen. Die jeweilige Inanspruchnahme liegt dabei im eigenen Verantwortungsbereich. Die Tutor:innen stehen für fachliche Rücksprachen zur Themenwahl einerseits sowie für formale und allgemeine Fragen zum wissenschaftlichen Arbeiten andererseits zur Verfügung. Eine Abnahme von Gliederungen, Textteilen oder –entwürfen durch die Tutor:innen ist hierbei jedoch nicht vorgesehen, da die eigenständige Erstellung Teil der zu erbringenden Prüfungsleistung ist und in die Gesamtbewertung einfließt. Es werden jedoch Hinweise zu Gliederungsentwürfen gegeben, um den Einstieg in die Strukturierung einer wissenschaftlichen Arbeit zu erleichtern.