

Projektarbeit im Rahmen der Vorlesung Statistik I - Aufgaben

Das Projekt geht über die Analyse von Flugdaten kommerzieller Flüge innerhalb der USA, die ursprünglich für den Zeitraum Oktober 1987 bis April 2008 vorliegen. Es werden für das Projekt konkret die Flugdaten größerer Flughäfen für das Jahr 2007 betrachtet, wobei jeder Gruppe die Daten eines einzelnen Abflug-Flughafen zugeteilt werden.

Führen Sie eine explorative Datenanalyse Ihres Datensatzes durch.

I. Spezifische Aufgaben:

- I.1. Welcher Prozentsatz der Flüge mit Abflugszeiten morgens zwischen 8 und 9 Uhr hatten mehr als 30 Minuten Verspätung?
- I.2. War die durchschnittliche Verspätung der verspäteten Flüge höher während der Woche oder am Wochenende?
- I.3. Betrachten Sie ArrTime gegen DepTime. Was fällt auf und woran liegt das? Kann man dieses Problem beheben?
- I.4. Betrachten Sie die gestrichenen Flüge. Gibt es Auffälligkeiten bzgl. des Zielflughafens, Monats oder Tags? Ist es hilfreich, den Grund der Streichung (CancellationCode) einzubeziehen?
- I.5. Welche fünf Airlines starten am häufigsten von Ihrem Flughafen? Mit welcher Airline sollte man fliegen, um möglichst wenig Verspätung zu haben?
- I.6. Geben Sie für alle Flugrouten die durchschnittliche Geschwindigkeit an. Gibt es signifikante Unterschiede? Sind Flüge, die verspätet gestartet sind, schneller als andere?

Weitere allgemeinere Fragestellungen, die untersucht werden könnten, sind die nachfolgenden.

II. Allgemeine Aufgaben:

- II.1. Auf welche Gründe sind Verspätungen eher zurückzuführen? Welche Rolle spielen hierbei Wettereinflüsse?
- II.2. Untersuchen Sie die verschiedenen "Delay" Arten untereinander. Verwenden Sie zur Unterstützung Ihrer Argumentation Graphiken und statistische Methoden.
- II.3. Welche anderen Variablen haben einen Einfluss auf Verspätungen?
- II.4. Fassen Sie eine Hauptaussage Ihrer Arbeit in einer geeigneten Graphik zusammen.

Alle Datensätze sind unterschiedlich, also vergewissern Sie sich, dass Sie den richtigen Datensatz für Ihre Projektgruppe von der Lehrveranstaltungs-Internetseite herunterladen. Bei Fragen und Problemen wenden Sie sich per E-mail an Ihren Betreuer! Informationen zu den Datensätzen findet man unter stat-computing.org/dataexpo/2009/ ; insbesondere eine Liste der Variablen unter stat-computing.org/dataexpo/2009/the-data.html .