Arbeitsauftrag: Big Data – Deutsche Bahn

1. **Was ist das Ziel von Davids Hobbyprojekt?**

Die Deutsche Bahn zu Vorratsdatenspeichern:

BahnMining:

Um zu checken ob die Daten der DB wahr sind.

1. **Wie viele Halte hat Davids Datensatz umfasst?**

25 Millionen

1. **Was ist die offizielle Definition von Pünktlichkeit bei der deutschen Bahn?**

Pünktlich = Verspätung < 6 Minuten

1. **Welcher Bahnhof hat die höchste Anzahl von Fernverkehrszügen pro Tag?**

Köln Hbf

1. **Welcher Monat hat die höchste Verspätungsquote im Jahr 2019?**

Juli wegen Hitze (heißester Tag des Jahres)

1. **Welche zwei Faktoren beeinflussen die Verspätung eines Zuges am meisten?**

Zugart (ICE/IC/EC)

Klimaanlage zu heiß/kalt

1. **Was ist der Unterschied zwischen der absoluten und der relativen Verspätung eines Zuges?**

Die absolute Verspätung beschreibt die Differenz zwischen der tatsächlichen Ankunfts- oder Abfahrtszeit eines Zuges und der planmäßigen Zeit.

Die relative Verspätung hingegen bezieht die geplante Reisezeit mit ein. Sie wird als Prozentsatz der planmäßigen Reisezeit berechnet.

1. **Wie hat David die Verspätungsdaten der deutschen Bahn gesammelt?**

Indem er Netzwerkanfragen an die Server der DB macht um die Pläne herunterzuladen und das möglichst minimiert.

1. **Welche Software hat er verwendet, um die Daten zu analysieren und zu visualisieren?**

Analysieren: Sehr wahrscheinlich in Javascript oder Python, da in diesen Sprachen die Fahrplan API verfügbar ist.

Visualisieren: Powerpoint

1. **Welche Methode hat er angewendet, um die Zuglänge aus den Daten abzuleiten?**

Indem er den Abfahrtzeitpunkt der ersten Station nimmt und den Ankunftszeitpunkt der letzten Station und sich damit die Zuglänge/Fahrtdauer berechnet.

1. **Wie hat er die Verspätungsquote eines Zuges berechnet?**

(Anzahl an Verspätungen + Anzahl an Ausfällen)/Insgesamte Anzahl an Stopps = Verspätungsquote in %

1. **Wie hat er die Verspätung eines Zuges in Abhängigkeit von der Zugart dargestellt?**

In einem Balkendiagramm und einer Art Blasendiagramm.

1. **Wie hat er die Verspätung eines Zuges in Abhängigkeit von der Tageszeit dargestellt?**

In einem Balkendiagramm.

1. **Wie hat er die Verspätung eines Zuges in Abhängigkeit von der Zuglänge dargestellt?**

Auch in einem Balkendiagramm. Man konnte sehen wie die Pünktlichkeit abstieg je länger der Zug fuhr.