

Bericht der Gruppe Ufo

Eine Analyse der Abhängigkeit der Wetters auf vermeintliche Ufo-Sichtungen.

Lars Thomsen, Uzeyir Mammadov

Martin-Luther-Universität Halle-Wittenberg

[lars.thomsen][uzeyir.mammadov]@student.uni-halle.de



Abbildung 1: Some large rubber duck.

ABSTRACT

KEYWORDS

Übung „Big Data Analytics“, Sommersemester 2021, Ufo

1 EINLEITUNG

Auf den ersten Blick klingen vermeintliche Ufo-Sichtungen immer nach an den Haaren herbeigezogenen Erfindungen und Lügengeschichten. Doch wie viel Wahrheit steckt in diesen Sichtungen, und lassen sich gegebenenfalls Muster in diesen erkennen?

Im Zuge unseres Projektes haben wir uns folgende Forschungsfrage gestellt: Existieren Abhängigkeiten zwischen vermeintlichen Ufo-Sichtungen, und welchen Einfluss nimmt das zum Zeitpunkt der Sichtung herrschende Wetter?

Der Bericht gliedert sich wie folgt: Kapitel 2 beschreibt die verwendeten Daten und Datensätze. In Kapitel 3 wird die Forschungsfrage und ihre Bearbeitung vorgestellt. Kapitel 4 diskutiert die Ergebnisse unserer Frage. Im Schlusskapitel geben wir unser Fazit und runden den Bericht ab.

2 DATEN

Die verwendeten Datensätze stammen aus einer Datenbank vom „National UFO Reporting Center“ (NUFORC). In diese Datenbank kann jede Person ihre vermeintliche Ufo-Sichtung – entweder online über ein Formular oder per Telefon – eintragen. Des Weiteren senden gewisse Messstationen (z.B. MOAfyxcc) auffällige Messdaten automatisch an die Datenbank. Um offensichtlichen Fakes entgegenzuwirken werden die eingereichten Daten vor der monatlichen Veröffentlichung von den Betreibern des Portals überprüft.

Jeder Eintrag des Datensatzes besteht aus dem Datum und der Uhrzeit der Sichtung, der Stadt sowie dem Kürzel des dazugehörigen Bundesstaates (ausschließlich für die

USA), einer klassifizierten Beschreibung der Form des Ufos und gegebenenfalls einer kurzen Zusammenfassung und Beschreibung aus der Sicht des Einsenders.

Für diese Daten haben wir eine einfache Datenbank erstellt, um auch offline auf diese zugreifen zu können. Um die Daten einfacher verarbeiten zu können, bereiten wir diese während der Abfrage von unserer Datenbank auf und treffen eine Vorauswahl an brauchbaren Daten. Die Kriterien für die Vorauswahl wurden durch stichprobenartige Abfragen festgelegt. Dem entsprechend werden Einträge von vorhin beschriebenen externen Messstationen, Einträge mit fehlenden Inhalten oder „?“ als Inhalt nicht betrachtet. Der Datentyp aller Attribute ist der Einfachheit halber nur string. Zu einem späteren Zeitpunkt werden Datum und Uhrzeit in ein datetime-Format überführt. Mit dem Wissen aus den stichprobenartigen Abfragen legen wir und auf die zwei gängigsten Formate fest – 'm/d/y H:M' und 'm/d/y'. Andere in dem Datensatz vorkommenden Formate oder von den Einsendern eigenständige Angaben beachten wir nicht. Unsere aufbereiteten Daten werden letztendlich durch das Tripel `datetime`, `city` und `state` beschrieben.

3 DIE ERSTEN VERSUCHE

Describe how you worked on the first question you address.

Dies hier ist ein Blindtext zum Testen von Textausgaben. Wer diesen Text liest, ist selbst schuld. $\sin^2(\alpha) + \cos^2(\beta) = 1$. Der Text gibt lediglich den Grauwert der Schrift an $E = mc^2$. Ist das wirklich so? Ist es gleichgültig, ob ich schreibe: Dies ist ein Blindtext oder Huardest gefburn? Kjift – mitnichten! Ein Blindtext bietet mir wichtige Informationen. $\sqrt[n]{a} \cdot \sqrt[n]{b} = \sqrt[n]{ab}$. An ihm messe ich die Lesbarkeit einer Schrift, ihre Anmutung, wie harmonisch die Figuren zueinander stehen und prüfe, wie breit oder schmal sie läuft. $\frac{\sqrt[n]{a}}{\sqrt[n]{b}} = \sqrt[n]{\frac{a}{b}}$. Ein Blindtext sollte möglichst viele verschiedene Buchstaben enthalten und in der Originalsprache gesetzt sein. $a \sqrt[n]{b} = \sqrt[n]{a^n b}$. Er muß keinen Sinn ergeben, sollte aber lesbar sein. $d\Omega = \sin \vartheta d\vartheta d\varphi$. Fremdsprachige Texte wie Lorem ipsum dienen nicht dem eigentlichen Zweck, da sie eine falsche Anmutung vermitteln.

4 FORSCHUNGSFRAGE

Describe how you worked on the second question you address.

Dies hier ist ein Blindtext zum Testen von Textausgaben. Wer diesen Text liest, ist selbst schuld. $\sin^2(\alpha) + \cos^2(\beta) = 1$. Der Text gibt lediglich den Grauwert der Schrift an $E = mc^2$. Ist das wirklich so? Ist es gleichgültig, ob ich schreibe: Dies ist ein Blindtext oder Huardest gefburn? Kjift – mitnichten! Ein Blindtext bietet mir wichtige Informationen. $\sqrt[n]{a} \cdot \sqrt[n]{b} = \sqrt[n]{ab}$. An ihm messe ich die Lesbarkeit einer Schrift, ihre Anmutung, wie harmonisch die Figuren zueinander stehen und prüfe, wie breit oder schmal sie läuft. $\frac{\sqrt[n]{a}}{\sqrt[n]{b}} = \sqrt[n]{\frac{a}{b}}$. Ein

Tabelle 1: Kennzahlen des verwendeten Datensatzes.

Datensätze	Anzahl
Gesamt	96 924
Davon einzigartige Orte	25 234
Orte mit Wetterdaten, gesamt	2 020
Davon mit Sonnenminuten pro Stunde	26
Davon mit Sonnenminuten pro Tag	116
Davon mit condition codes	1 886

Blindtext sollte möglichst viele verschiedene Buchstaben enthalten und in der Originalsprache gesetzt sein. $a \sqrt[n]{b} = \sqrt[n]{a^n b}$. Er muß keinen Sinn ergeben, sollte aber lesbar sein. $d\Omega = \sin \theta d\theta d\phi$. Fremdsprachige Texte wie Lorem ipsum dienen nicht dem eigentlichen Zweck, da sie eine falsche Anmutung vermitteln.

5 EVALUATION

Some evaluation section if appropriate. You might want to refer to some table with results in this section (e.g., to Table 1).

6 FAZIT

The introduction in less detail. Summarize story in retrospective, give outlook on possible next steps. Semi-technical

...