**Aufgabe 2.1: Beschreibung des Datensatzes „Drama“**

* **ID**: Die ID dient der schnellen Zuweisung und bewirkt dadurch, dass doppelte Einträge beim „label“ auch eindeutig identifizierbar sind. Gäbe es z.B. den Namen „Gertrude“ in einem Werk zweimal, wüsste der Suchende im ersten Moment nicht, ob es sich dann auch um die gesuchte Person handelt. Dies verhindert die Zuweisung einer ID.
* **label**: Wie schon bei der ID erwähnt, können sich im label identische Begriffe in Folgezeilen wiederholen. Das label identifiziert eine Person im Theaterstück, macht dies jedoch nicht eindeutig wie die ID. Würde man eine Theaterstückauflistung von Shakespeares Dramen manchen und nicht wie in diesem Fall die Personenliste eines einzelnen Stückes, so wären die

*labels* nur durch die ID schnell zuordenbar, außer man würde müsste durch zusätzliche Suche mehr Informationen zur Person herausfinden, wie das Stück, in dem sie vorkommt, ihre Bedeutung für die Handlung etc. Dennoch ist eine ID nur eine Zahl und der Name das, was die Person eigentlich ausmacht, beide, sowohl *label* als auch ID sind in einer Auflistung direkt miteinander verbunden, daher sollte auch nicht nur das *label* ohne ID-Zuweisung verwendet werden.

* **gender**: Ebenso wie ID und Name, lässt sich auch das Geschlecht in der Regel eindeutig zuweisen. Bei einem Geist ist jedoch zusätzliches Wissen nötig, das sich nur durch Nachlesen der Personenbeschreibung oder der Kenntnis eines Werkes erschließen lässt, ansonsten müsste der Geist dem Geschlecht neutral oder divers entsprechen. Das Geschlecht ist vor allem bei der Betrachtung von Geschlechterrollen wichtig oder zur Untersuchung wie viele Frauen und Männer im Theaterstück vorkommen.
* **role**: Die Rollenzuweisung lässt sich nur durch Kenntnis des Stückes erschließen. Es muss bekannt sein, ob es sich um den Protagonisten, Antagonisten, Geliebte/-r, oder einen sonstigen Charakter handelt. Die Rolle lässt sich gut mit dem Geschlecht verknüpfen. Dadurch wird z.B. die Rolle der Frau in einer jeweiligen Epoche hervorgehoben, wie oft beispielsweise eine Frau als Protagonistin, Antagonistin oder Geliebte auftritt im Vergleich zum Mann. Auch kann die Rolle einen Bezug zum *label* haben, wenn das Label z.B. einen Adelstitel wie König/in besitzt, selbstverständlich lässt sich so ein Bezug auch zwischen Geschlecht und Titel herstellen.
* **importance**: Die Bedeutsamkeit einer Person des Theaterstücks lässt sich auf zwei Arten festlegen: subjektiv oder objektiv. Eine subjektive Festlegung seitens des Lesers ist nur durch dessen Wissen über das Stück möglich. Der Leser bewertet aus seiner Sicht einen Charakter nach dessen Handeln im gesamten Werk und legt die entsprechende Wichtigkeit der Rolle nach den Gesichtspunkten – in diesem Fall *primär, sekundär, weniger wichtig* – fest.
* **per\_mes\_sps**: Dies Eintrag zählt das Auftreten einer Person in Sprechakten und kann eng mit deren Bedeutsamkeit verbunden werden. Wie zuvor unter *importance* beschrieben, lassen sich zwei Arten der Bedeutsamkeit einer Person festlegen, nämlich subjektiv oder objektiv. Der objektive Weg die Wichtigkeit einer Person zu finden, ist zu überprüfen, wie oft die Person vorkommt. Man könnte beispielsweise aus allen Werten einen Durchschnittswert nehmen, diesen als Grenze festlegen und daraufhin in gleichmäßigen Etappen prüfen, ob die Person überdurchschnittlich oder unterdurchschnittlich oft auftritt. Eine weitere Möglichkeit wäre den Maximalwert heranzuziehen, diesen durch die Anzahl der Bedeutsamkeitsstufen zu teilen und so bestimmte Grenzen festzulegen. Trotzdem ist eine objektive Herangehensweise nicht immer zielführend. So hat in Shakespeares Hamlet der Geist eine tragende Rolle, die nur durch einen kurzen Auftritt ohne Sprechakte den gesamten Verlauf der Handlung bestimmt. Es ist in vielen Fällen somit zusätzliches Wissen notwendig. Doch nicht nur *importance* besitzt einen Bezug zu der Anzahl der Sprechakte. So lassen sich über die Anzahl der Sprechakte auch die Protagonisten und Antagonisten anhand untersuchen, z.B. wie oft der jeweilige Charakter auftritt und etwas äußert. Auch *gender* spielt eine Rolle betrachtet man folgende Fragen: Wie oft kommen Frauen zu insgesamt Wort? Wie viele Sprechakte bekommen Frauen im Durchschnitt in Werken der jeweiligen Epoche?

**Aufgabe 2.2: Numpy und Pandas**

S. Datei *aufgabe02\_02.ipynb* im Anhang des Uploads!!!