Pig Use Case - The Show Data Analysis Part

# Problem Description

We have with us historical data of The Show guests from 1999 to 2004.

**Data Categorization:**

* **Year:** The year an episode was aired.
* **GoogleKnowledge\_Occupation:** Their occupation or office, according to Google’s Knowledge Graph. On the other hand, if they are not in there, how Stewart introduced them on the program.
* **Show:** Air date of an episode. Not unique, as some shows had more than one guest.
* **Group:** A larger group designation for the occupation. For instance, US senators, US presidents, and former presidents are all under “politicians.”
* **Raw\_Guest\_List:** The person or list of people who appeared on the show, according to Wikipedia. The GoogleKnowlege\_Occupation only refers to one of them in a given row.

**Problem Statement 1:**

**Find the top five kinds of GoogleKnowlege\_Occupation people who were guests in the show, in a particular time period.**

**Sample Output:**

(actor,28)

(actress,20)

(comedian,4)

(television actress,3)

(singer,2)

**Problem Statement 2:**

**Find out the number of politicians who came each year.**

**Sample Output:**

(2014,13)

(2000,13)

(2013,11)

(2002,8)

(2001,3)

(1999,2)