

# Data Engineering Midterm Presentation

---

GROUP 9

The logo for IPVS, consisting of the letters "IPVS" in a white, bold, sans-serif font, set against a solid black square background.

**IPVS**

# Goal

---

1. For the musicians that went to the WH, how many have worked on cinema?
2. What musicians have acted and worked on the soundtrack for the same movie?
3. Look for musicians/actors from a same movie that went to WH and the movie has or is about the WH.

# Goal 1

---

Find musician that went to the White House and worked on cinema.

Used last.fm and IMDb APIs and the White House dataset.

Limitations due to the APIs connections.

```
# df stands for WhiteHouse dataframe
for index, row in df.iterrows():
    name = df.at[index, 'NAMEFIRST'] + " " + df.at[index, 'NAMELAST']
    if (not(name in searched)):
        url = # url Last.fm API
        response = requests.get(url)
        searched.append(name)
        ret = artist_info(response.content)
        if ret is not None:
            ret_imdb = i.search_person(name)
            for p in range(len(ret_imdb)):
                if (str(ret_imdb[p]).upper() == str(name).upper()):
                    print 'FOUND IN IMDB ' + name
                    results.append(ret)
                    break
```

# Goal 1

From the top 10 of the first 3000 White House registers:

James Carr ('2010-05-19')

Carlos Montoya ('2010-05-19')



## James Carr

Singer

James Carr, was an American rhythm and blues and soul singer. Born to a Baptist preacher's family in Coahoma, Mississippi, Carr began singing in church and was performing in gospel groups and making ... [Wikipedia](#)

**Born:** June 13, 1942, [Coahoma](#), [Mississippi](#), [United States](#)

**Died:** January 7, 2001, [Memphis](#), [Tennessee](#), [United States](#)

**Genres:** [Rhythm and blues](#), [Soul music](#)

# Goal 1

From the top 10 of the first 3000 White House registers:

James Carr ('2010-05-19')

Carlos Montoya ('2010-05-19')



## Carlos Montoya

Guitarist



Carlos García Montoya in Madrid, Spain, was a prominent flamenco guitarist and a founder of the modern-day popular flamenco style of music. A few of his video recordings can still be found on YouTube.  
[Wikipedia](#)

**Born:** December 13, 1903, Madrid, Spain

**Died:** March 3, 1993, Wainscott, New York, United States

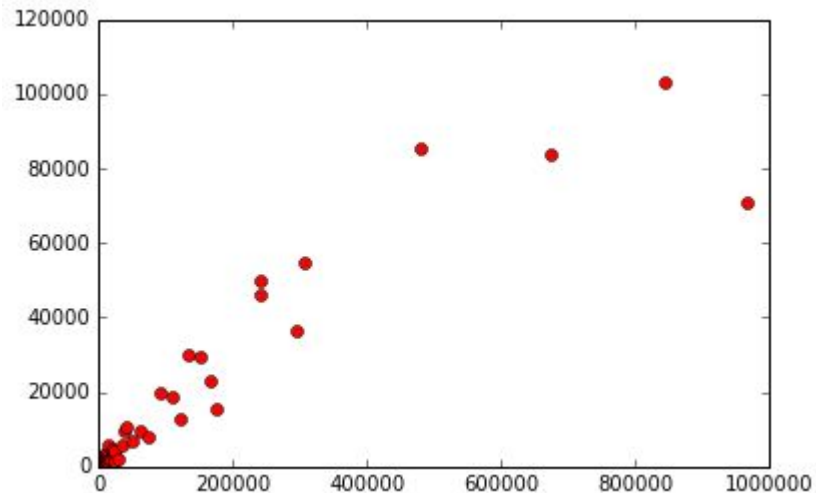
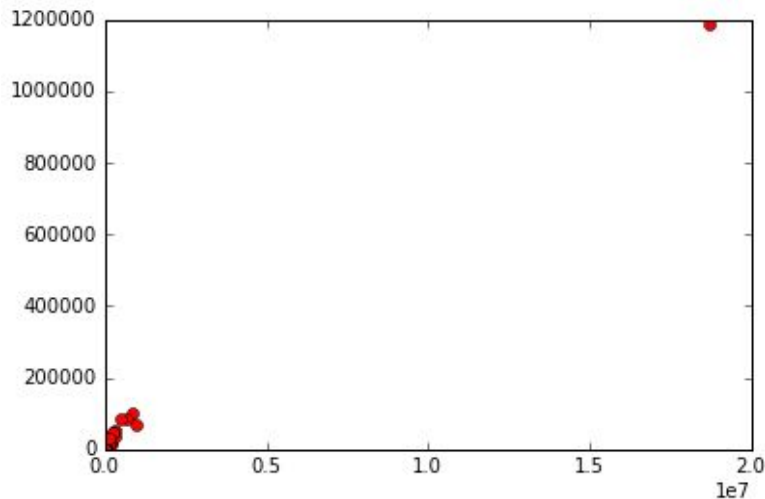
**Genre:** Flamenco

**Albums:** [Spanish Guitar Magic](#), [The Art of Flamenco](#), [more](#)

**TV shows:** [Showcase '68](#)

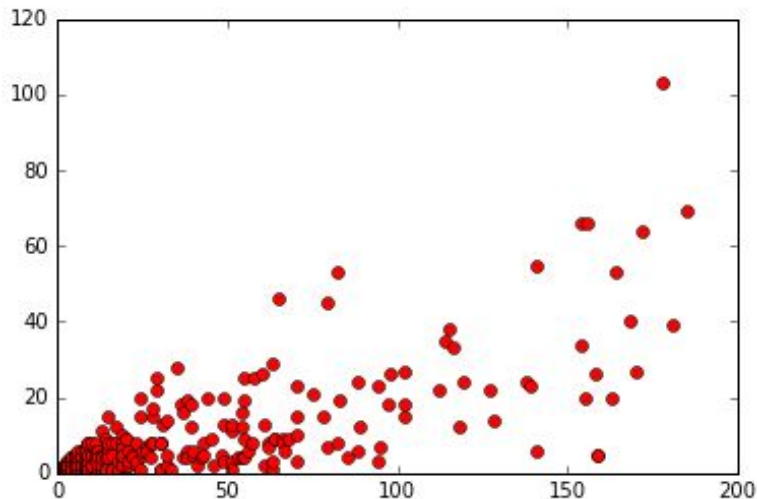
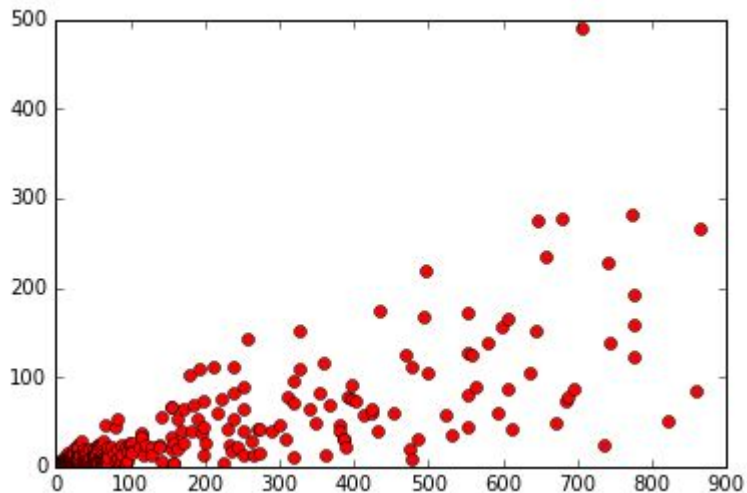
# Goal 1

Relation *listeners* : *playcount* for each musician that is an artist and went to the White House.



# Goal 1

Relation *listeners* : *playcount* for each musician that is an artist and went to the White House.



# Goal 2

---

What musicians have acted and worked on the soundtrack for the same movie?

## **What we did:**

Convert from .list files to csv files

Got performer information from soundtrack.list file

**Has problems....**



# Goal 2

---

Problem, Cast Information of Movie:

- From csv files
  - Three separate files: title, name and cast\_info
  - title(movies) and name(actors) don't have the corresponding ids
  - Can't get cast information from csv files
- From imdbpy
  - Some films don't have cast information

# Goal 3

---

Idea similar to Goal 1.

Needs only to find movies about the White House in the IMDb dataset.

Not yet done due to the same problems found in Goal 2.