CIS 550 Project
Milestone 1

HU CHENCHEN 11021747
XIANG CHEN 66770481
ZHONG WEN 54397279
DHWANI KAPOOR 84152362

Background and Initiatives

This project is going to use the Summer Olympics dataset and others to make predictions on several interesting things such as the number of medals and explore various aspects such as social networks of athletes.

Primary goals:

1. Show general statistics of the Olympic games with respect to national ranking, best players, composition of medals as well as the changes between different years.
2. Build models to predict the number of medals, the distribution of medals, winner and performance of certain nationality/gender/athlete.

Questions:

1. What does the social profile of a particular athlete look like?
2. What can we know about the development or emerging trend of certain sports?
3. From data, is it possible to describe the impacts of political, financial or cultural events on the performance of Olympics?
4. What is the economic impact of hosting the Olympic Games?
5. What are participation and performance disparities among different countries, areas or races?
6. How do the performance in Olympic competitions contribute to an athlete's career?

Methodology

Some features we might use to build prediction models:
1. The influence of historical events/macroeconomics/technology development.
2. The influence of geographical/political/financial features on the performance of a country.
3. The social profile of a certain athlete (social networks, world ranking, news etc.).

Technologies and APIs:

1. Node.js, also some modules in Node.js like Express.
2. D3.js
3. AngularJS
4. HTML/CSS/jQuery
5. Facebook, Twitter, Google API
6. AWS
    a. Relational Databases Service (SQL)
    b. DynamoDB(NoSQL)

Timeline

October 10th - Develop an initial idea
October 31st - Improve the initial idea and make it more specific and realizable
November 14th - Prepare the needed data and design the database
November 23rd - Develop the web
December 12th - Complete the whole project and prepare for the demonstration

Division of Responsibility

Responsibility:
Data extraction and populating the database (data cleaning, formatting and entity resolution) -
Hu Chenchen
Schema design - Xiang Chen
Web design - Zhong Wen
Prediction model - Dhwani

Experimental validation:
Latency - Dhwani, Xiang Chen
Concurrent requests - Hu Chenchen, Zhong Wen

Milestones:

Milestone 2:
1. Relational Schema design
    ● ER diagram
    ● SQL DDL (normalized)

- Description of NoSQL component(optional)

Milestone 3:
1. Data extraction and populating the database (data cleaning, formatting and entity resolution)
2. Translating Milestone 1 questions/amended questions to SQL queries
3. AWS/ Oracle database on Amazon RDS, text file creation (full JDBC/SQLPLUS) connect string

Milestone 4:
1. Application development(basic features, advanced features)
2. Source code and list of features

Milestone 5:
1. Final demo and report development
2. Experimental validation (concurrent requests?, latency?)