



## **PROJECT 3 – PROPOSAL SUBMISSION**

**COURSE TITLE**

**CS 363L  
Database Systems [Lab]**

**SEMESTER**

**Sixth  
Spring 2022**

**SUBMITTED BY**

**2019-CE-3 (Sundas Noreen)  
2019-CE-4 (Saba)**

**SUBMITTED TO**

**Ms. Darakhshan Abdul Ghaffar**

**DATE OF SUBMISSION**

**March 27, 2022**

---

**Department of Computer Engineering  
University of Engineering and Technology, Lahore**

## **PROJECT 3 – PROPOSAL**

### **Dataset to be Used:**

Hacker News

### **About the Dataset:**

This dataset contains all stories and comments from Hacker News from its launch in 2006. Each story contains a story id, the author that made the post, when it was written, and the number of points the story received. Hacker News is a social news website focusing on computer science and entrepreneurship. It is run by Paul Graham's investment fund and startup incubator, Y Combinator. In general, content that can be submitted is defined as "anything that gratifies one's intellectual curiosity".

### **Content:**

Each story contains a story ID, the author that made the post, when it was written, and the number of points the story received.

### **Size:**

The dataset is available on big query with the name bigquery-public-data.hacker\_news and is of 19.2132 GB.

### **Questions:**

1. Recent studies have found that many forums tend to be dominated by a very small fraction of users. Is this true of Hacker News?
2. Hacker News has received complaints that the site is biased towards Y Combinator startups. Do the data support this?
3. Is the amount of coverage by Hacker News predictive of a startup's success?