**Lab assignment 1**

**CIS 492**

**Fabio Hinojosa Jimenez**

**2861151**

**No extra credit**

Table of Contents

[Part 3 3](#_Toc162787721)

[Preprocessed Data 3](#_Toc162787722)

[Calculating Proximity 3](#_Toc162787723)

[Simple Matching 3](#_Toc162787724)

[Extended Jaccard Similarity 4](#_Toc162787725)

[Cosine Similarity 4](#_Toc162787726)

[Results 4](#_Toc162787727)

# Part 3

## Preprocessed Data

A screenshot of a computer

Description automatically generated

## Calculating Proximity

We first start by reading the data, we select only the numeric values, and select only the rows with ‘11000’, ‘11001’, and ‘11012’ from CustomerKey.

A screenshot of a computer code

Description automatically generated

### Simple Matching

For simple matching we sum all of the values that are equal and then we divide by the length of the row.

A black and blue text

Description automatically generated

### Extended Jaccard Similarity

By using a logical and to identify the elements that intersect with each other, same way we use the logical or to identify the union of the elements, and following the equation we divide the intersection and union to get the answer.

A screenshot of a computer code

Description automatically generated

### Cosine Similarity

Following the cosine formula we first get the dot product of the two vectors, then using numpy we calculate the Euclidean length of each vector, and to finalize we divide the results.

A screenshot of a computer code

Description automatically generated

## Results

Finally we print the results.

A screenshot of a computer

Description automatically generated