Section A

```
In [2]:  import time
         start_time = time.process_time()

         with open('kosarak.dat') as f:
           lines = f.read().split( '\n')

         articles = []
         users = []

         for line in lines:
           stripped_line = line.rstrip()
           if stripped_line != "":
             line_components = stripped_line.split(' ')
             user_articles = set()

             for article in line_components:
               if article != "":
                 article_num = int(article)
                 if article_num not in articles:
                   articles.append (article_num)
                 user_articles.add(article_num)
             users.append(sorted(user_articles))

         article_strings = [("@attribute news%d {0,1}" % (num - 1)) for num in articles]
         user_strings = ["{%s}" % ",".join ("%d 1" % (num - 1) for num in user) for user in users]
         all_lines = ["@relation kosarak"] + article_strings + ["@data"] + user_strings
         with open("kosarak.arff", "w") as f:
           for line in all_lines:
             f.write(line + "\n")

         print(f"Time elapsed: {time.process_time() - start_time} seconds")
```
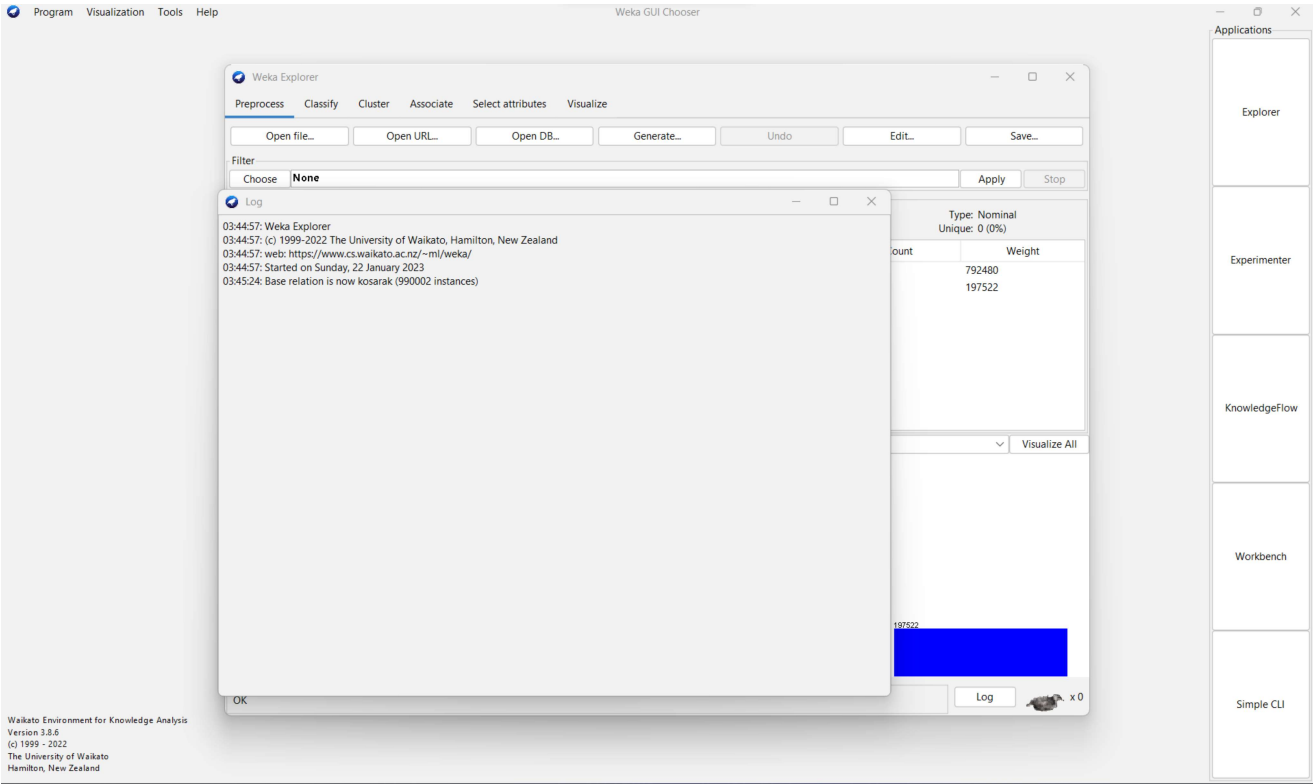
```
Time elapsed: 237.79842923700002 seconds
```
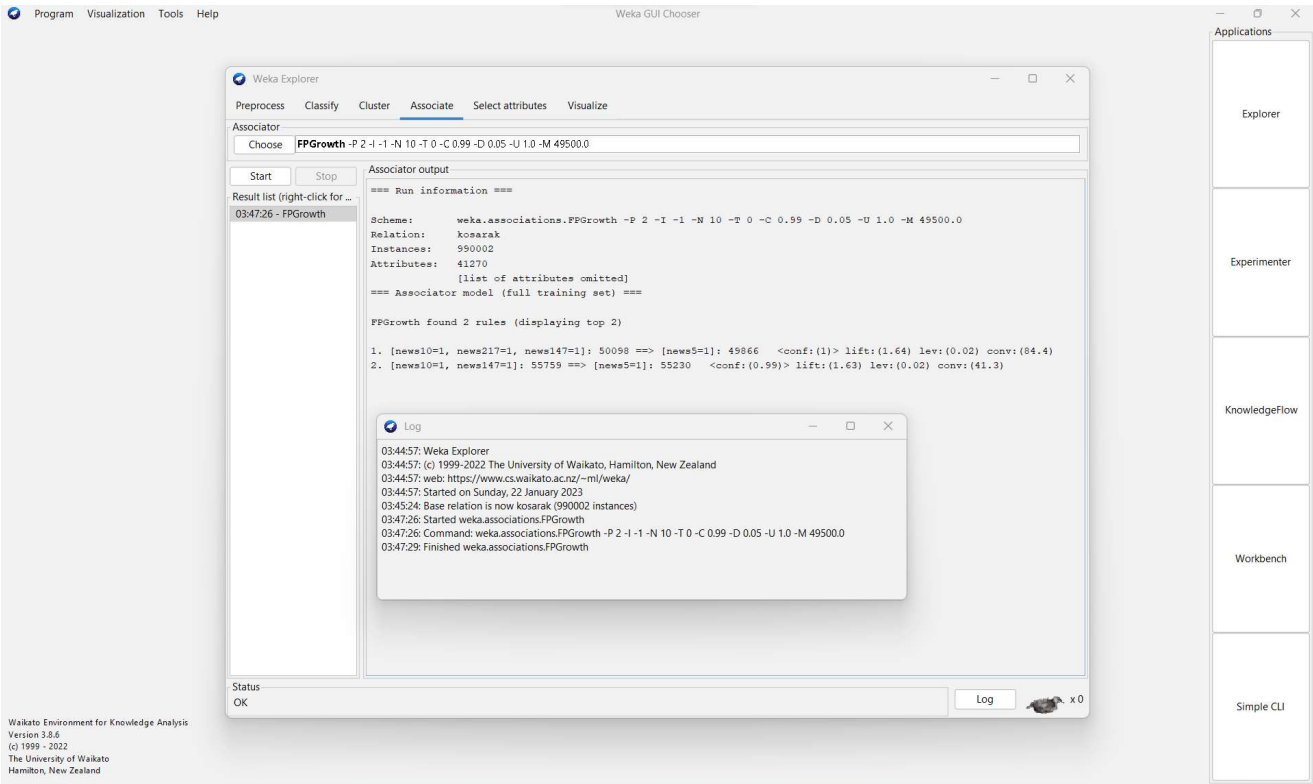
Section B

The time taken for converting kosark.dat to kosark.arff file is 237.79 seconds.

Section C

The time taken to load the file in Weka was 27 seconds.

Section D

Type *Markdown* and LaTeX: α^2

Section E

For the first time it took 3 seconds.

For the second time it took 7 seconds.

The third time it took 6 seconds.

The fourth time it took 4 seconds.

The fifth time took 5 seconds.

The Average time taken is 5 seconds which is way less than the time taken to convert the dataset and then load into Weka.