Homework1

October 7, 2024

1 NAME: Ramya Chowdary Patchala

2 Homework 1 - IST 652

Complete the following tasks in the cells provided. If you need additional cells, please add them. To submit your solution, save it first (in JupyterHub), then download the notebook as a .ipynb file to your local computer system and submit it via Blackboard. (Use JupyterHub's "File" menu and the "Download" option in it to download the notebook file)

1. Create a string called *myfirstname* and store your first name in it

[2]:
```
myfirstname = "Ramya"
```

2. Create a string called *mylastname* and store your last name in it

[3]:
```
mylastname = "Patchala"
```

3. Create a string called *myname* that concatenates the strings *myfirstname* and *mylastname* that you created previously

[4]:
```
myname = myfirstname +" " + mylastname
print(myname)
```

```
Ramya Patchala
```

4. Use the appropriate methods seen in the class session to do the following on the string *myname*:

4a. Print out the number of characters (i.e. the length) that make up *myname*

[5]:
```
# len() gives no.of chars in string
len(myname)
```

[5]:
```
14
```

4b. Derive a substring composed of the letters in positions 2 to 5 from *myname*. The characters in positions 2 and 5 must be included

[7]:
```
# Slicing from the index 1 which is 2nd letter to 5th index - obtaining␣
 ↪substring
myname[1:5]
```

[7]:
```
'amya'
```

4c. Choose a letter that is present in myname and replace it with the character X

[8]:
```
# Replace function can replace char in string
myname = myname.replace('a','X')
print(myname)
```

```
RXmyX PXtchXlX
```

5. Create a list called *WeekDays* that contains the names of the day of the week.

[9]:
```
WeekDays = ['Monday', 'Tuesday', 'Wednesday', 'Thursday', 'Friday', 'Saturday',␣
↪'Sunday']
```

6. Create a dictionary that uses the course numbers of the courses you are taking this semester as the keys of the dictionary and for each key associate as its value the day of the week in which you have that class. As an example for IST 652, the entry would be *652 : Wednesday*

[10]:
```
courses = {652:'Monday', 645:'Tuesday', 688:'Thursday'}
courses?
```

```
Type:        dict
String form: {652: 'Monday', 645: 'Tuesday', 688: 'Thursday'}
Length:      3
Docstring:
dict() -> new empty dictionary
dict(mapping) -> new dictionary initialized from a mapping object's
    (key, value) pairs
dict(iterable) -> new dictionary initialized as if via:
    d = {}
    for k, v in iterable:
        d[k] = v
dict(**kwargs) -> new dictionary initialized with the name=value pairs
    in the keyword argument list.  For example:  dict(one=1, two=2)
```

7. Do some research on how to use Markdown commands and write the text of the first three sections of your resume (example: education, objectives, past experience, etc.).

You can use the following resource to learn about Markdown: - https://medium.com/analytics-vidhya/the-ultimate-markdown-guide-for-jupyter-notebook-d5e5abf728fd - Short video: https://www.youtube.com/watch?v=uVLzL5E-YBM - Long video/tutorial: https://www.youtube.com/watch?v=mTIifW_LU5s

Jupyternotebook Tutorial

- https://www.youtube.com/watch?v=DKiI6NfSIe8&t=711s

3 RAMYA CHOWDARY PATCHALA

(315) 901-9074 • rpatchal@syr.edu • LinkedIn • GitHub

3.1 EDUCATION

3.1.1 Syracuse University, School of Information Studies, Syracuse, NY

M.S. Applied Data Science | Aug 2023 - May 2025
GPA: 4.0

Relevant Coursework:
- Applied Machine Learning - Natural Language Processing - Data Warehouse - Introduction to Data Science - Data Administration Concepts and Database Management - Quantitative Reasoning for Data Science - Scripting for Data Analysis - Building Human-Centered AI Applications (LLM) - Managing Information Systems Projects

3.1.2 Jawaharlal Nehru Technological University, Kakinada, Andhra Pradesh, India

B.Tech, Computer Science and Engineering | Jun 2015 - May 2019

Relevant Coursework:
- Database Management Systems - Data Warehousing and Mining - Design and Analysis of Algorithms - Advanced Data Structures - Probability and Statistics - Cloud Computing - Human-Computer Interaction - Hadoop and Big Data

---

3.2 WORK EXPERIENCE

3.2.1 Data Analyst, Sidearm Sports, United States

May 2024 - Aug 2024
- Developed and deployed new websites based on stakeholder requirements, utilizing web scraping techniques to gather and integrate data from client websites. - Collaborated with various departments to manage web development projects, ensuring alignment with stakeholder goals and contributing to program and project campaigns. - Drafted, designed, updated, and revitalized Excel lists and spreadsheets to track project progress and ensure data accuracy for website development tasks.

3.2.2 Data Scientist, SG Analytics, Hyderabad, India

Sep 2022 - Aug 2023
- Developed a name matching engine for NBC Universal using Machine Learning and Natural Language Processing techniques, improving data matching accuracy by 20%. - Developed applications using JavaScript and SQL in Retool to sync data between Peacock, Nielsen, and IMDb, reducing processing time by 30%. - Designed data pipelines in Databricks using PySpark and Python to merge IMDb data with other client-specific sources, improving efficiency by 25%. - Migrated Retool applications from AWS DynamoDB to Snowflake, reducing query response time by 30%.

3.2.3 Data Engineer, Tata Consultancy Services, Hyderabad, India

Jul 2019 - Sep 2022
- Maintained a data pipeline with 99.8% uptime, ensuring continuous data collection from 8 main sources for Western Union. Used Apache Spark for real-time processing and Python for automation. - Built automated ETL pipelines using Google Cloud Dataflow, improving the data integration process for BigQuery storage. - Integrated Tableau with GCP data infrastructure for real-time updates and enabled data-driven strategies for biller performance management. - Optimized complex SQL queries, improving their performance by 50%.

3.3 PROJECTS

3.3.1 Document Question Answering System

- Developed and fine-tuned a document QA system using Llama and LangChain, enabling users to upload documents and ask context-aware questions.
- Integrated Streamlit for a user-friendly interface and Chroma for efficient similarity search and retrieval.

3.3.2 Image Caption Generator

- Developed an advanced image captioning system using BLIP, achieving a BLEU score of 0.72 and generating detailed, context-aware descriptions of images.
- Utilized Vision Transformer and BLIP models for optimal performance.