UCSC

SCS2211 - LABORATORY II
R Lab Practical Sheet - 09

Instructions

- **Do the Activities in a R markdown file**
- **File name should be <Index number>.rmd (Eg: 2000000.rmd)and upload to the given link.**
- **Any form of plagiarism or collusion is not allowed**

Jupyter Notebook

What is Jupyter Notebook?

Jupyter Notebook is an open-source, web-based interactive computing platform that allows users to create and share documents containing live code, equations, visualizations, and narrative text. It supports various programming languages, with Python being the most commonly used. Jupyter Notebook is widely used in data analysis, scientific computing, machine learning, and educational settings due to its ability to combine code execution, rich text formatting, and visualizations in a single document. Its interactive nature makes it a powerful tool for developing and presenting data-driven projects.

Installing Jupyter Notebook

- Install Python and Jupyter:
    - These commands are executed in the R console (inside RStudio or any R environment):
        - install.packages("reticulate")
        - reticulate::install_miniconda()
        - reticulate::conda_install("r-reticulate", "jupyter")
- Add R kernel to Jupyter:
    - install.packages("IRkernel")
    - IRkernel::installspec()

Launching Jupyter Notebook

- Start Jupyter Notebook:
    - jupyter notebook
- In the Jupyter interface, select New > R to create an R notebook.

Activity 01

Dataset : https://www.kaggle.com/sonukumari47/students-performance-in-exams

1. Download the dataset
2. Read the dataset to a variable
3. Change the column names to followings

Original	New
X	id
race.ethnicity	race
parental .level.of.education	edu
lunch	lunch
test.preparation.course	prep
math.percentage	math
reading.score.percentage	read
writing.score.percentage	write
sex	gen

4. Change the math, read and write variables to whole number (Multiply by 100)
5. Create a new attribute average (average of math,read and write)
6. Find the summary for each mark (math,read and write)
7. Find unique values for attributes
    a. race
    b. edu
    c. lunch
    d. prep
    e. gen
8. Find average mark for
    a. race
    b. edu
    c. lunch

d. prep
e. gen

Activity 02

Dataset : https://www.kaggle.com/dansbecker/melbourne-housing-snapshot

1. Download the dataset
2. Read the dataset to a variable
3. Print first few values of the dataset
4. Find number of missing values in each attribute
5. Find the mean value for “YearBuilt”

Activity 03

1. Create the following dataframe (emp_sal)

Emp ID	Dep	Basic	Allowances
11	Sales	25450	5200
12	HR	22500	4500
13	Sales	21000	3100
14	HR	23500	2600
15	Sales	15000	1800

2. Store net salary in new column named “net_sal”
3. Obtain net salaries of employees attached to sales department
4. Obtain employee IDs of employees whose net salary is above 25000
5. Obtain employee IDs of employees attached to HR Department whose net salary is below 25000