**Charotar University of Science and Technology [CHARUSAT]**

**Chandubhai S. Patel Institute of Technology [CSPIT]**

**U & P U. Patel Department of Computer Engineering**

**Lab Manual**

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Subject code | : | CE-350 | Semester | : | 6 | Academic Year | : | 2021 |
| Subject name | : | Data Warehouse and Data Mining | | | | | | |

**Practical – 2**

**Aim: Data pre-processing using pandas:**

* **Prerequisite**

S/W: - Anaconda and Jupyter Notebook

Little bit knowledge about the python programming and some theory about the data pre-processing.

* **Step to install software**

1: [Download the Anaconda installer](https://www.anaconda.com/download/#windows).



2: RECOMMENDED: [Verify data integrity with SHA-256](https://docs.anaconda.com/anaconda/install/hashes/). For more information on hashes, see [what about cryptographic hash verification?](https://conda.io/projects/conda/en/latest/user-guide/install/download.html#cryptographic-hash-verification)

3: Double click the installer to launch.

**Note**

To prevent permission errors, do not launch the installer from the [Favorites folder](https://docs.anaconda.com/anaconda/user-guide/troubleshooting/#distro-troubleshooting-favorites-folder).

**Note**

If you encounter issues during installation, temporarily disable your anti-virus software during install, then re-enable it after the installation concludes. If you installed for all users, uninstall Anaconda and re-install it for your user only and try again.

4: Click Next.

5: Read the licensing terms and click “I Agree”.

6: Select an install for “Just Me” unless you’re installing for all users (which requires Windows Administrator privileges) and click Next.

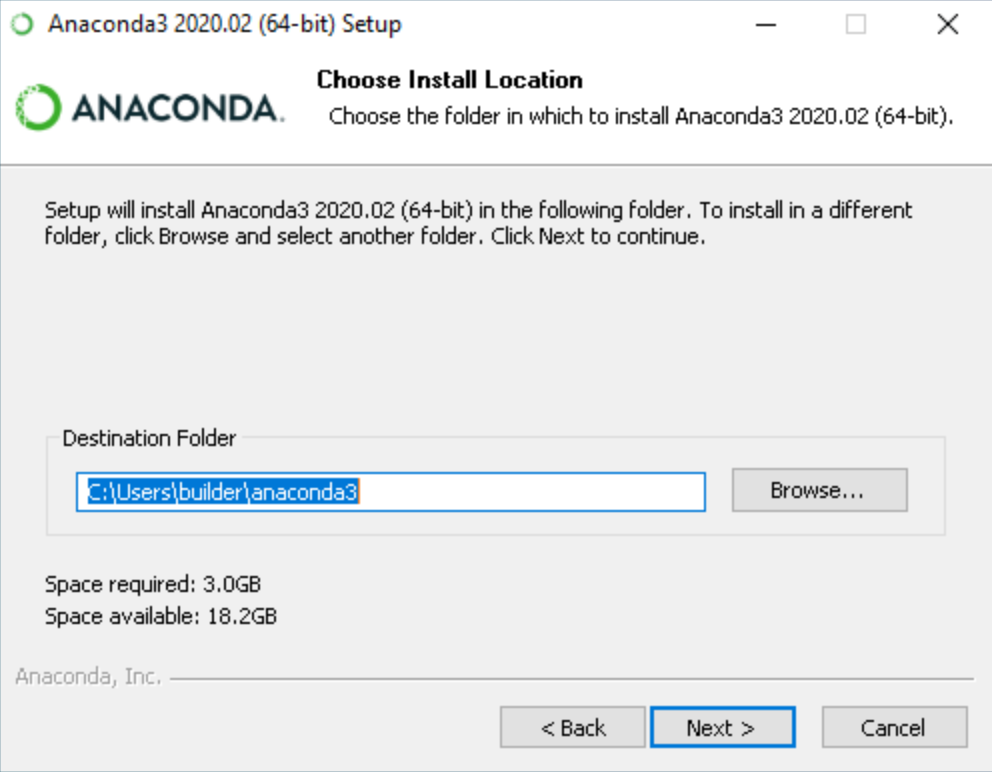
7: Select a destination folder to install Anaconda and click the Next button. See [FAQ](https://docs.anaconda.com/anaconda/user-guide/faq/#distribution-faq-windows-folder).

**Note**

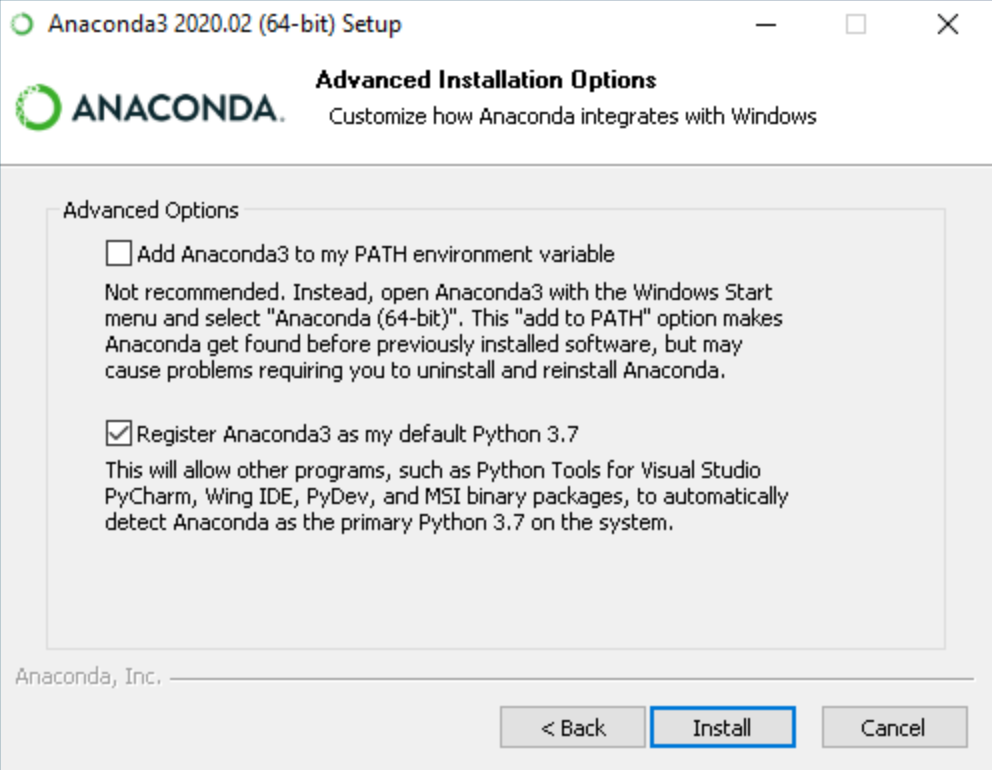
Install Anaconda to a directory path that does not contain spaces or unicode characters.

**Note**

Do not install as Administrator unless admin privileges are required.

[](https://docs.anaconda.com/_images/win-install-destination.png)

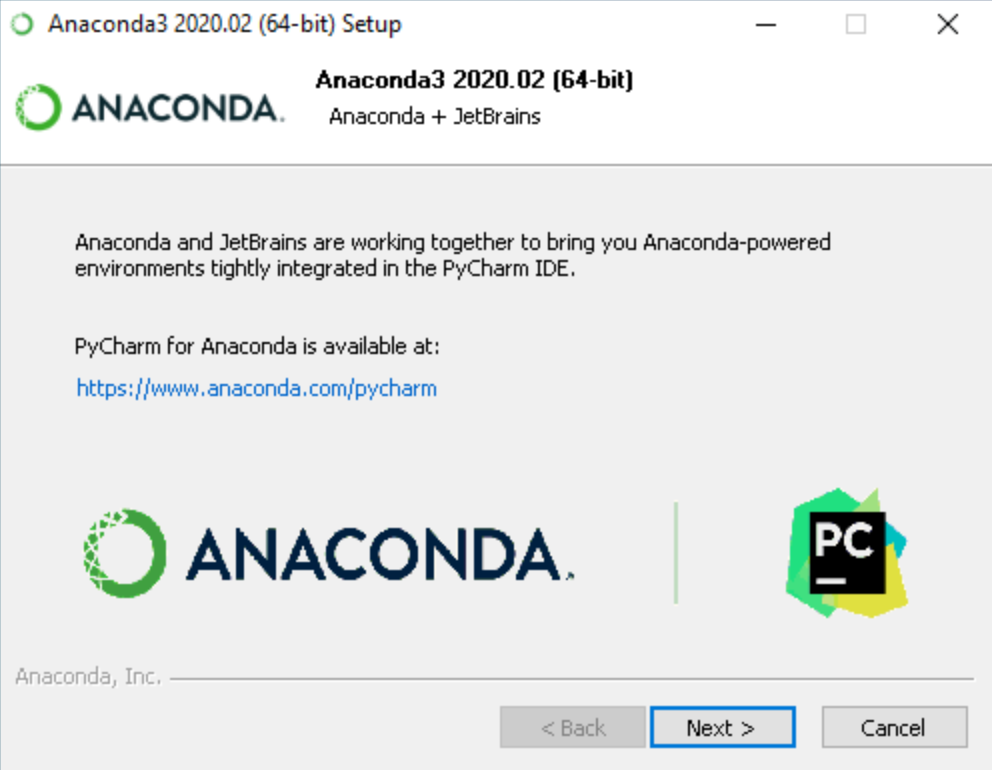
8: Choose whether to add Anaconda to your PATH environment variable. We recommend not adding Anaconda to the PATH environment variable, since this can interfere with other software. Instead, use Anaconda software by opening Anaconda Navigator or the Anaconda Prompt from the Start Menu.

[](https://docs.anaconda.com/_images/win-install-options.png)

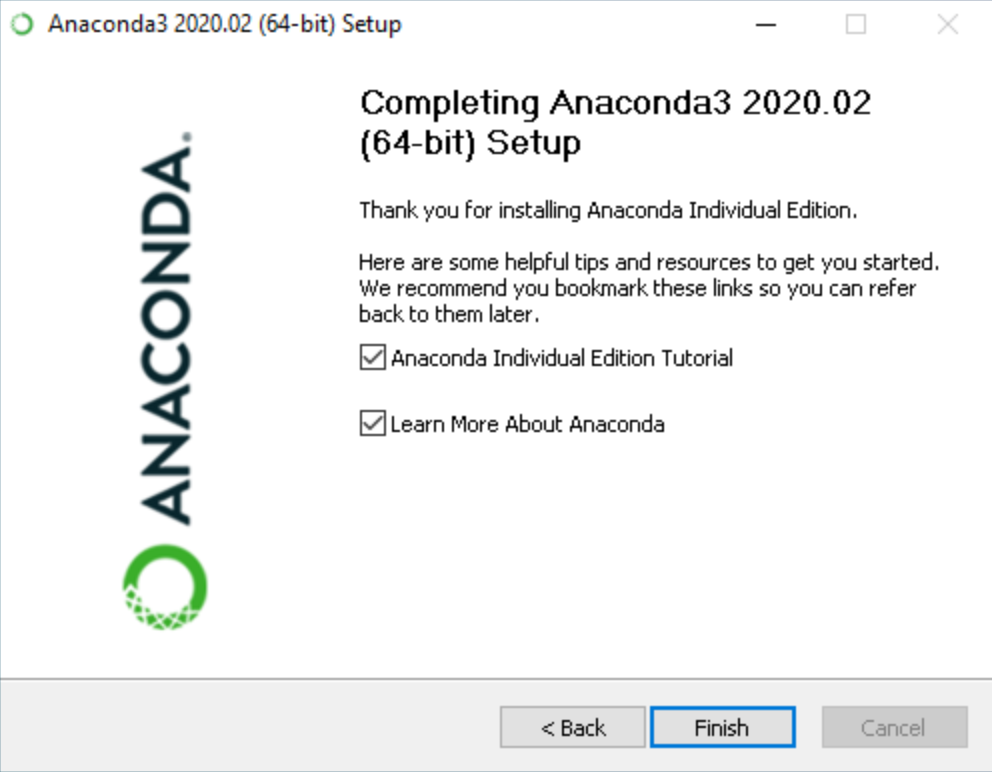
9: Choose whether to register Anaconda as your default Python. Unless you plan on installing and running multiple versions of Anaconda or multiple versions of Python, accept the default and leave this box checked.

10: Click the Install button. If you want to watch the packages Anaconda is installing, click Show Details.

11: Click the Next button.

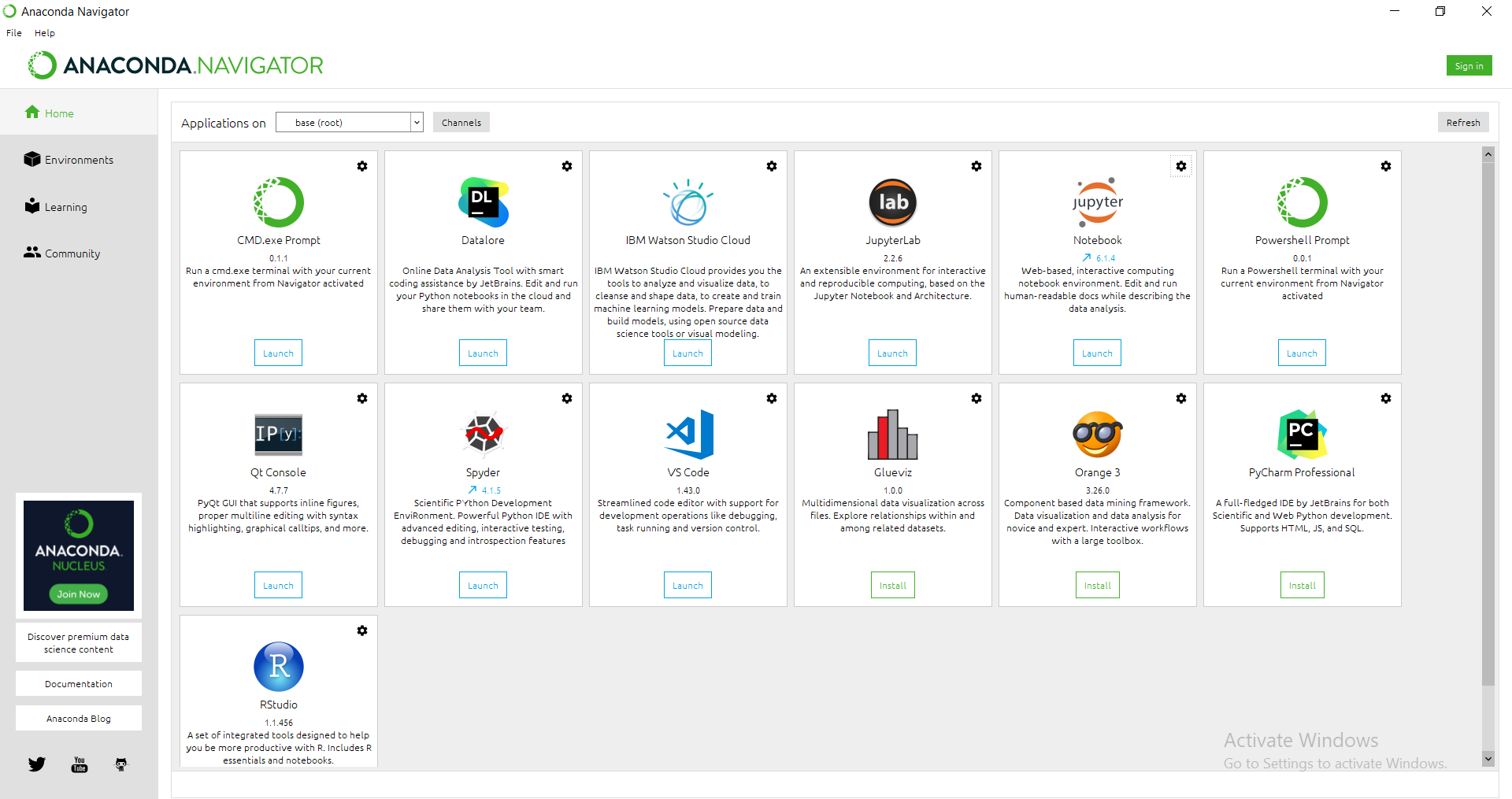
[](https://docs.anaconda.com/_images/win-install-pycharm.png)

12: After a successful installation you will see the “Thanks for installing Anaconda” dialog box:

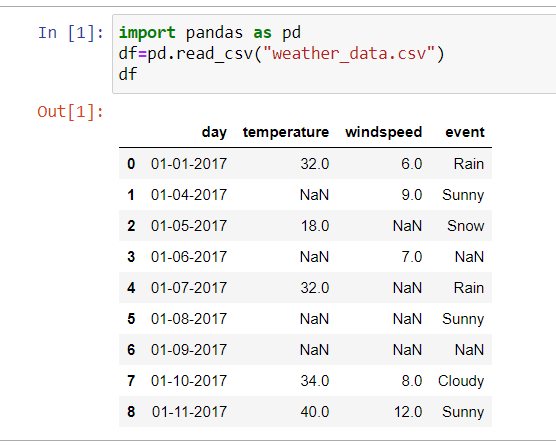
[](https://docs.anaconda.com/_images/win-install-complete.png)

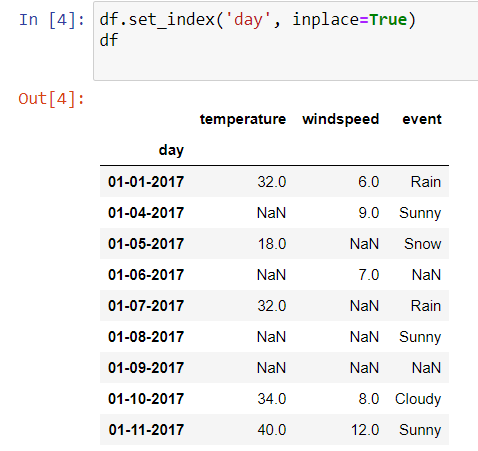
13: Click the Finish button.

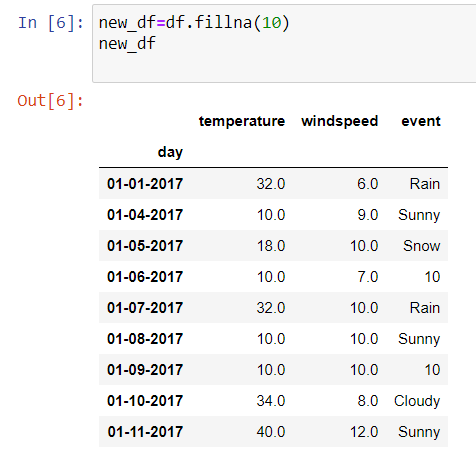
14: [Verify your installation](https://docs.anaconda.com/anaconda/install/verify-install/).

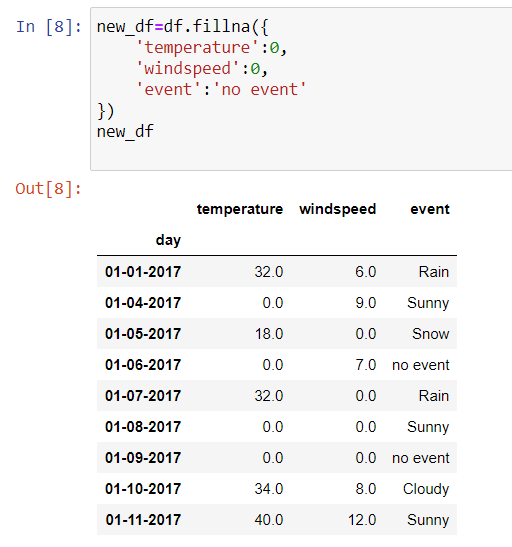


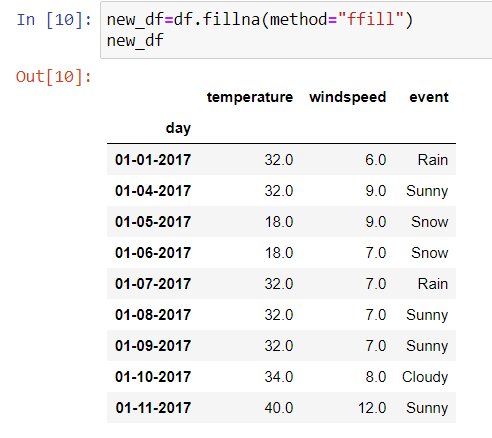
* **Screenshot of weather Data:**

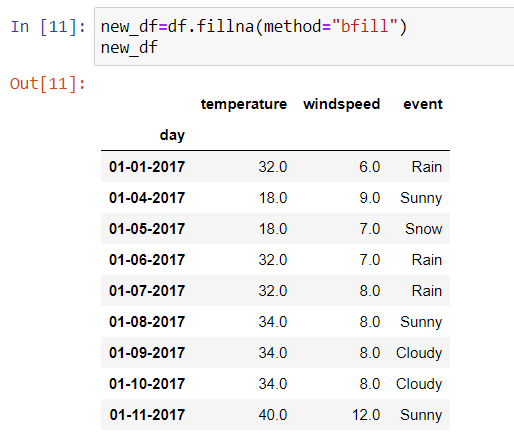


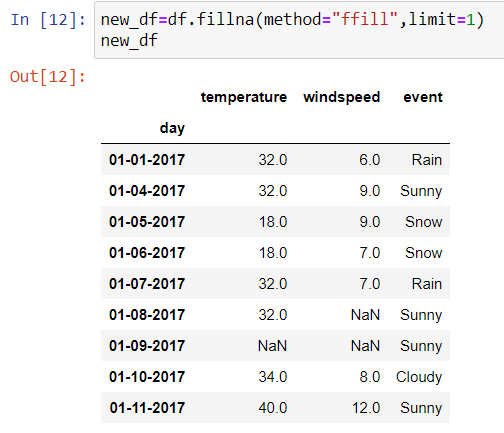


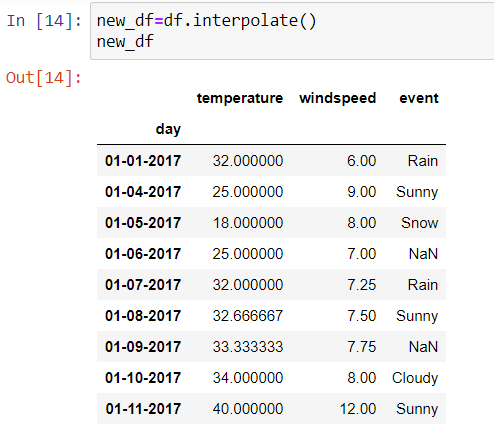


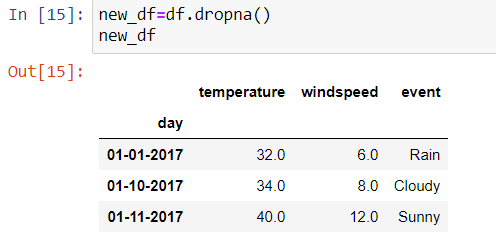


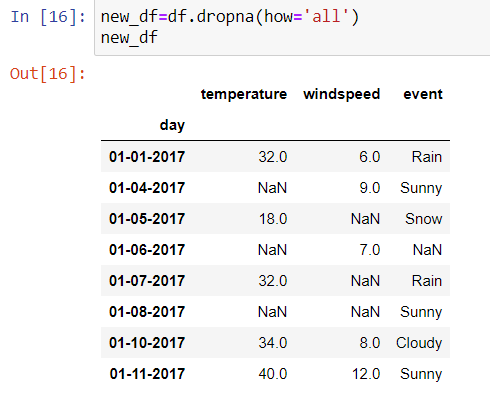


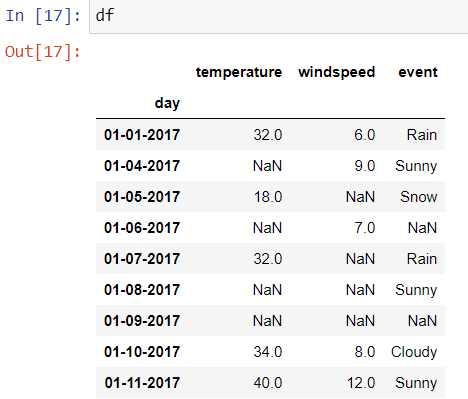


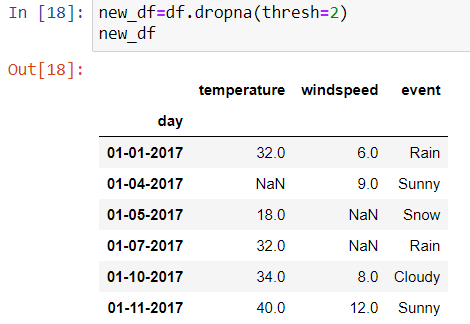


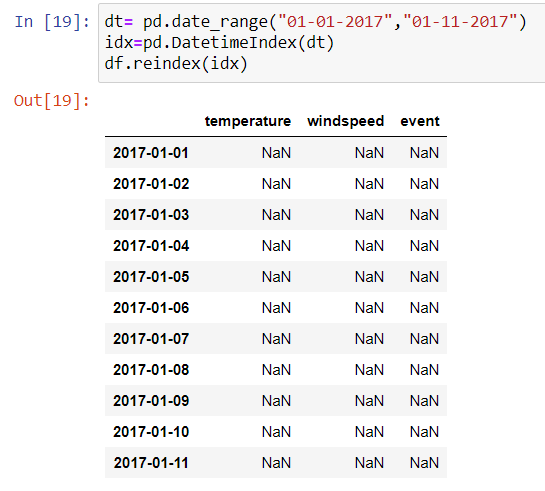




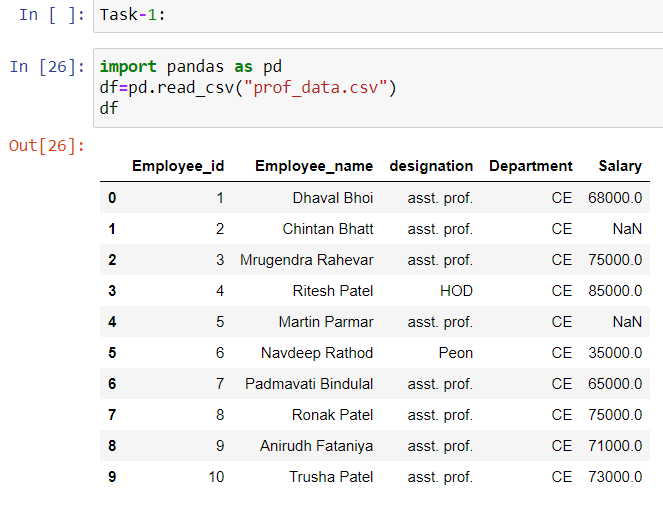


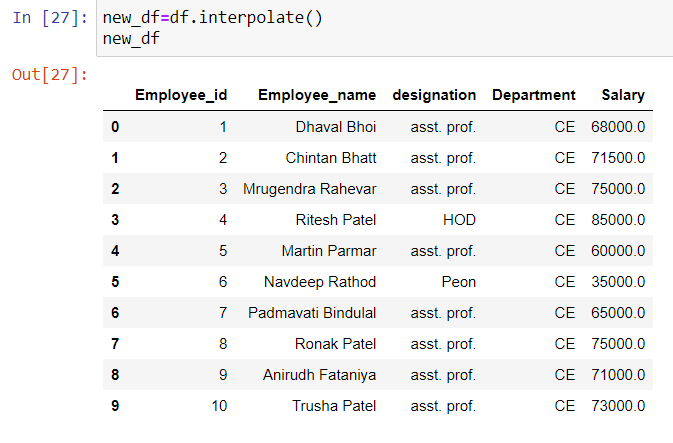






* **Screenshot of faculty Data**





* **Reference link of web material**

<https://www.ques10.com/p/9224/why-is-data-preprocessing-required-explain-the-dif/>

<https://docs.anaconda.com/anaconda/install/windows/>

<https://www.anaconda.com/products/individual>

* **5 Question/Answers**

**1: What is data pre-processing?**

The real world data’s are normally noise data so before organizing the data warehouse we need to preprocess the data.

**2: What is the discretization processes involved in data preprocessing?**

It reduces the number of values for a given continuous attribute by dividing the range of the attribute into intervals. Interval labels can then be used to replace actual data values.

**3: What is preprocessing technique?**

Data cleaning

Data integration

Data transformation

Data reduction

Define data cleaning

**4: why do we clean data?**

Having clean data will ultimately increase overall productivity and allow for the highest quality information in your decision-making. Benefits include: Removal of errors when multiple sources of data are at play. Fewer errors make for happier clients and less-frustrated employees.

**5: why is data quality important?**

Improved data quality leads to better decision-making across an organization. The more high-quality data you have, the more confidence you can have in your decisions. Good data decreases risk and can result in consistent improvements in results.

* **Conclusion**

From this practical we have learnt how to pre-process the data using the panda’s library, pre-analysis of data in order to transform them into a standard and normalized format. We are also learnt to deal with missing value, data formatting, data normalization, data standardisation, data binning etc.

|  |  |  |  |
| --- | --- | --- | --- |
| **Prepared By:** | Rajiv Kumar Gupta (18CE137) | **Date:** | 02-03-2021 |