File Format

- File formats are designed to store specific types of information, such as **CSV, XLSX** etc.
- The file format also tells the computer how to display or process its content. Common file formats, such as **CSV, XLSX, ZIP, TXT** etc.
- If you see your future as a data scientist so you must understand the different types of file format. Because data science is all about the data and it’s processing.
- If you don’t understand the file format so may be it’s quite complicated for you. Thus, it is mandatory for you to be aware of different file formats.

CSV

- **CSV:** the CSV stands for Comma-separated values. It uses comma to separate values. In CSV file each line is a data record and Each record consists of one or more then one data fields, the field is separated by commas.

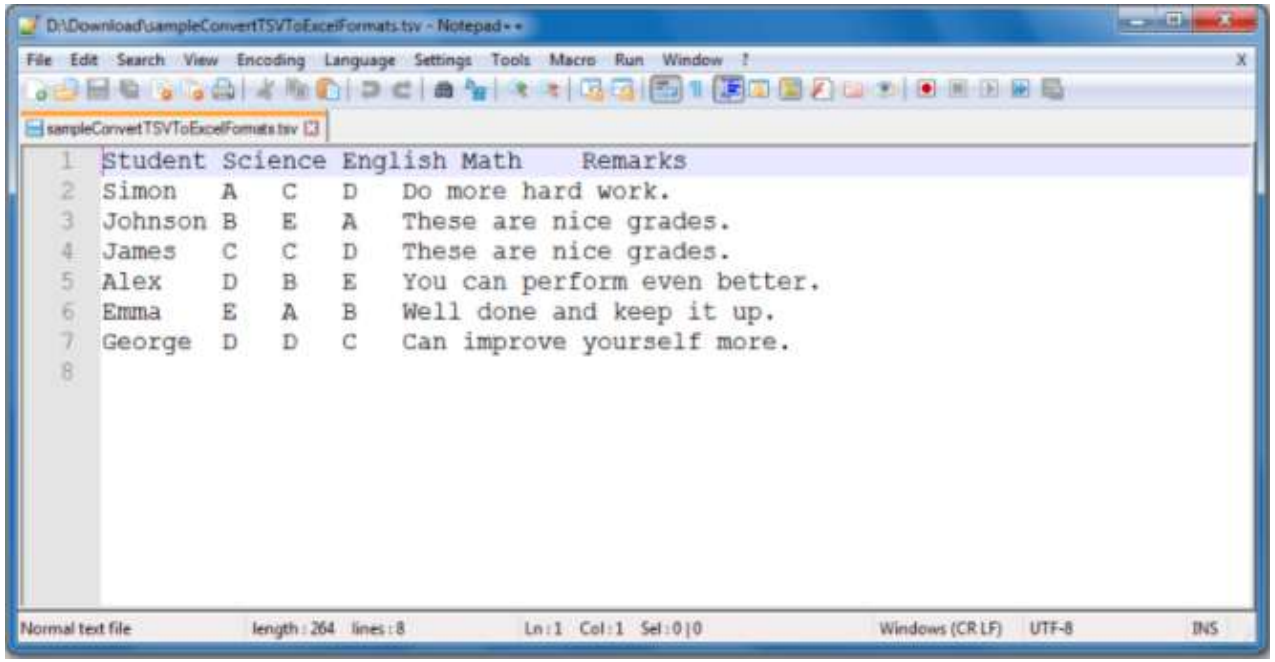
persons.csv - Notepad

File Edit Format View Help

```
Family Name,Given Name,VIAF ID
Ackersdijck,Willem Cornelis,17959345
Adelung,Friedrich von,22963658
Afzelius,Arvid August,49972119
Amerling,Karel,13331054
Anton,Karl Gottlob von,183632821
Arwidsson,Adolf Ivar,8184878
Asbjørnsen,Peter Christen,116587918
Attems,Heinrich,37665468
Atterbom,Per Daniel Amadeus,46819248
Balabin,Viktor Petrovich,44473845
Banks,Joseph,46830189
Beck,Friedrich,44338671
Becker,Reinhold von,42101066
Bernhart,Johann Baptist,69674335
Bertram,Johann,32890043
Bilderdijk,Willem,14882166
Boisserée,Sulpiz,7483155
Bopp,Franz,61614118
Borovský,Karel Havlíček,100277614
Bosković,Jovan,161354270
Buslaev,Fyodor,10074560
Cenowa,Florian Stanislaw,44466031
Chomiakov.Aleksei.66492873
```

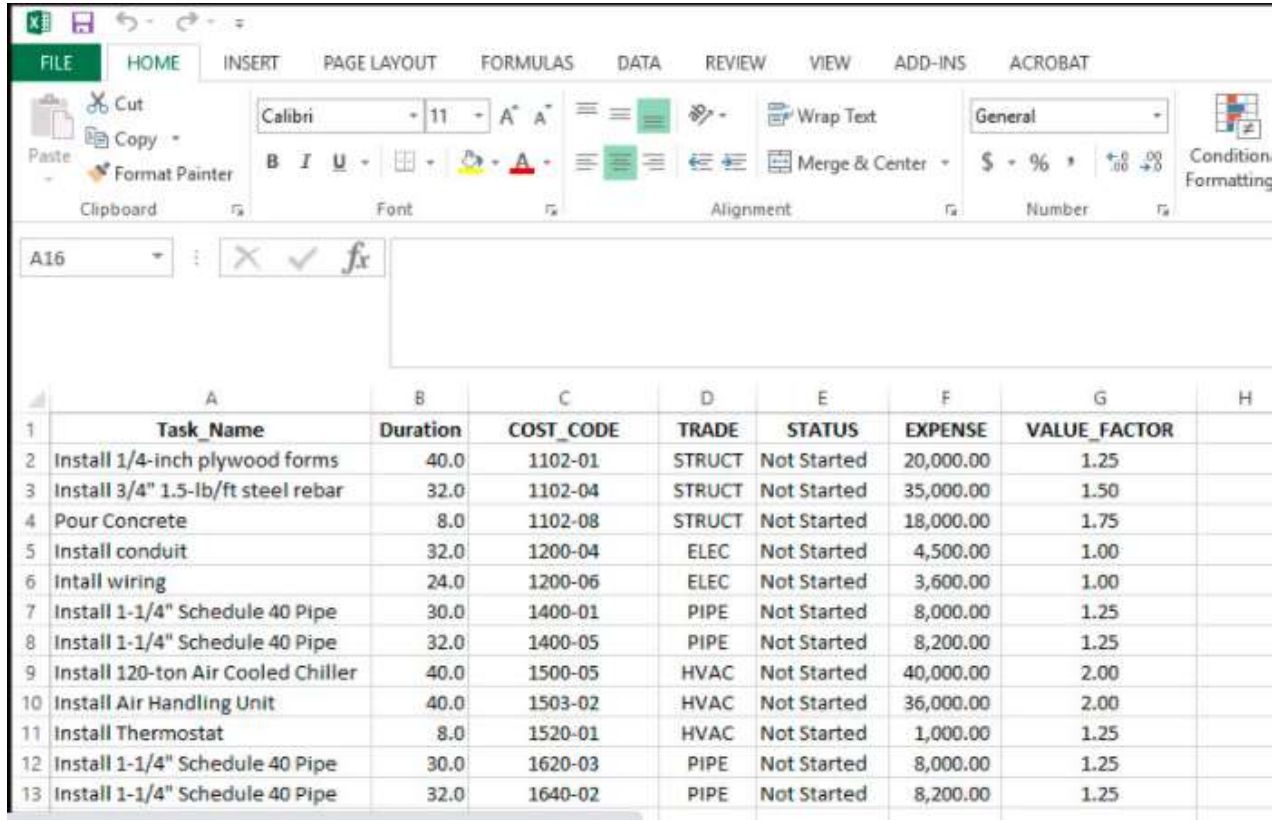
TSV

- Tab Separated Values.
- The TSV format is one of the most common formats for transferring data between applications and databases. It is an alternative format to .CSV.
- However, CSV files use commas to separate columns of data instead of tabs.
- TSV files are especially helpful for transferring data saved in a proprietary format into another program that does not support the format. For example, you can export email addresses in a spreadsheet application to a TSV file to upload it to an online emailing service. Or, you can export financial information from an online financial service to a TSV file, then import the file into a spreadsheet application.



XLSX

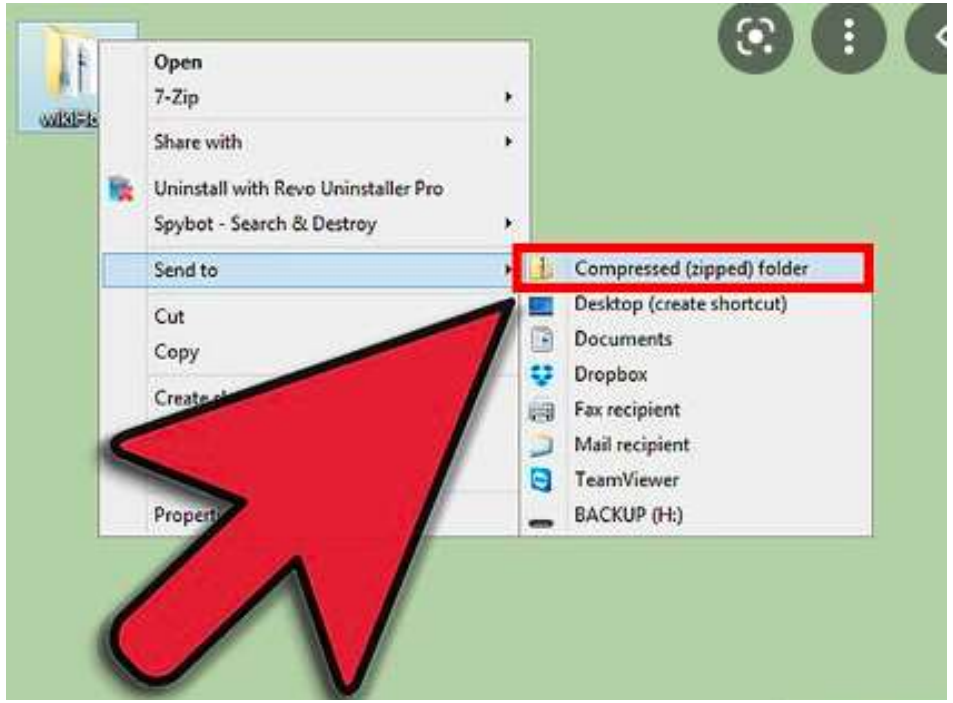
- **XLSX:** The XLSX file is Microsoft Excel Open XML Format Spreadsheet file. This is used to store any type of data but it’s mainly used to store financial data and to create mathematical models etc.



	A	B	C	D	E	F	G
1	Task_Name	Duration	COST_CODE	TRADE	STATUS	EXPENSE	VALUE_FACTOR
2	Install 1/4-inch plywood forms	40.0	1102-01	STRUCT	Not Started	20,000.00	1.25
3	Install 3/4" 1.5-lb/ft steel rebar	32.0	1102-04	STRUCT	Not Started	35,000.00	1.50
4	Pour Concrete	8.0	1102-08	STRUCT	Not Started	18,000.00	1.75
5	Install conduit	32.0	1200-04	ELEC	Not Started	4,500.00	1.00
6	Intall wiring	24.0	1200-06	ELEC	Not Started	3,600.00	1.00
7	Install 1-1/4" Schedule 40 Pipe	30.0	1400-01	PIPE	Not Started	8,000.00	1.25
8	Install 1-1/4" Schedule 40 Pipe	32.0	1400-05	PIPE	Not Started	8,200.00	1.25
9	Install 120-ton Air Cooled Chiller	40.0	1500-05	HVAC	Not Started	40,000.00	2.00
10	Install Air Handling Unit	40.0	1503-02	HVAC	Not Started	36,000.00	2.00
11	Install Thermostat	8.0	1520-01	HVAC	Not Started	1,000.00	1.25
12	Install 1-1/4" Schedule 40 Pipe	30.0	1620-03	PIPE	Not Started	8,000.00	1.25
13	Install 1-1/4" Schedule 40 Pipe	32.0	1640-02	PIPE	Not Started	8,200.00	1.25

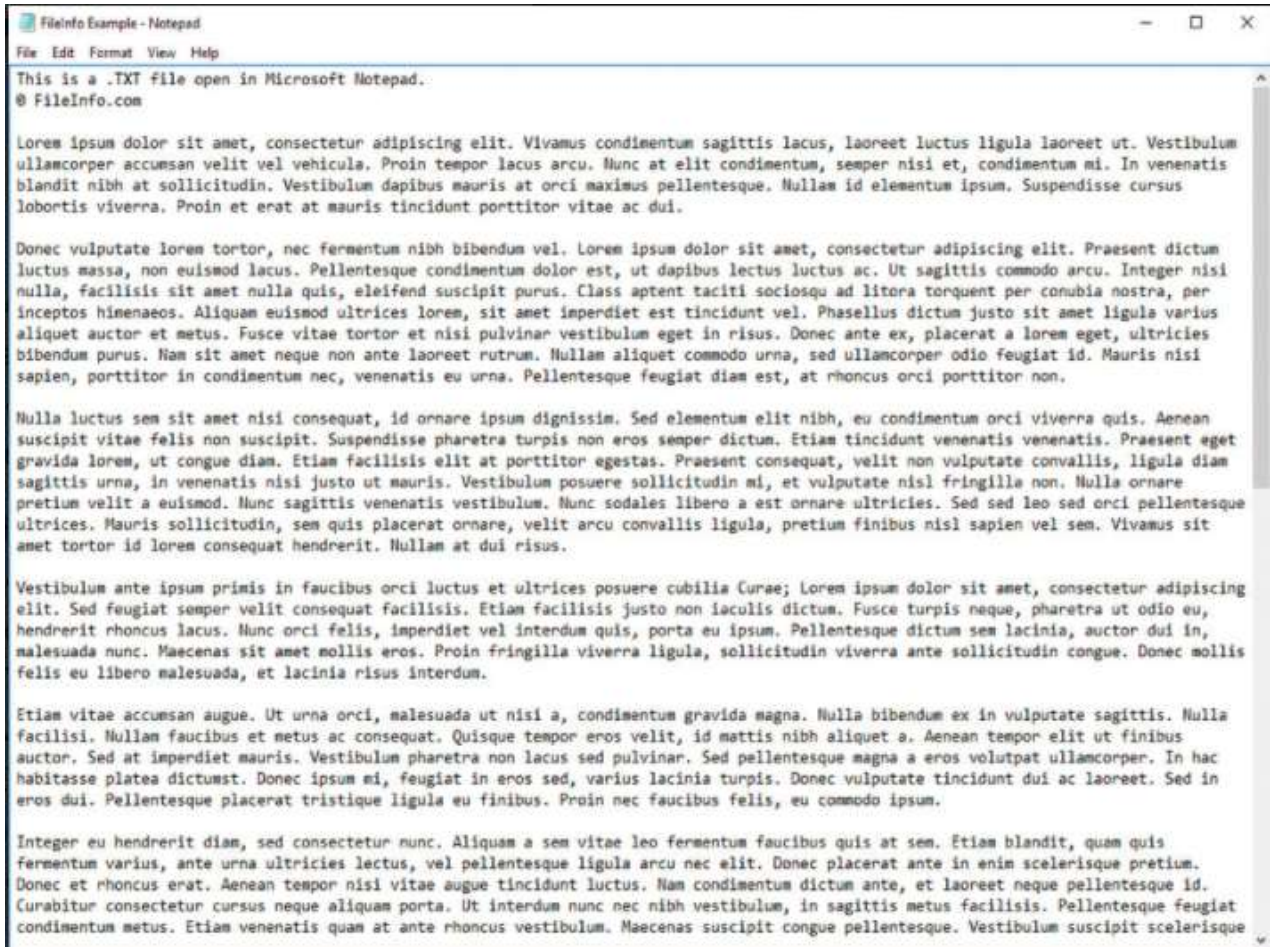
ZIP

- **ZIP:** ZIP files are used as data containers, they store one or more than one file in the compressed form. It widely used in internet After you downloaded ZIP file, you need to unpack its contents in order to use it.



TXT

- **TXT:** TXT files are useful for storing information in plain text with no special formatting beyond basic fonts and font styles. It is recognized by any text editing and other software programs.



FileInfo Example - Notepad

File Edit Format View Help

This is a .TXT file open in Microsoft Notepad.
© FileInfo.com

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Vivamus condimentum sagittis lacus, laoreet luctus ligula laoreet ut. Vestibulum ullamcorper accumsan velit vel vehicula. Proin tempor lacus arcu. Nunc at elit condimentum, semper nisi et, condimentum mi. In venenatis blandit nibh at sollicitudin. Vestibulum dapibus mauris at orci maximus pellentesque. Nullam id elementum ipsum. Suspendisse cursus lobortis viverra. Proin et erat at mauris tincidunt porttitor vitae ac dui.

Donec vulputate lorem tortor, nec fermentum nibh bibendum vel. Lorem ipsum dolor sit amet, consectetur adipiscing elit. Praesent dictum luctus massa, non euismod lacus. Pellentesque condimentum dolor est, ut dapibus lectus luctus ac. Ut sagittis commodo arcu. Integer nisi nulla, facilisis sit amet nulla quis, eleifend suscipit purus. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Aliquam euismod ultrices lorem, sit amet imperdiet est tincidunt vel. Phasellus dictum justo sit amet ligula varius aliquet auctor et metus. Fusce vitae tortor et nisi pulvinar vestibulum eget in risus. Donec ante ex, placerat a lorem eget, ultricies bibendum purus. Nam sit amet neque non ante laoreet rutrum. Nullam aliquet commodo urna, sed ullamcorper odio feugiat id. Mauris nisi sapien, porttitor in condimentum nec, venenatis eu urna. Pellentesque feugiat diam est, at rhoncus orci porttitor non.

Nulla luctus sem sit amet nisi consequat, id ornare ipsum dignissim. Sed elementum elit nibh, eu condimentum orci viverra quis. Aenean suscipit vitae felis non suscipit. Suspendisse pharetra turpis non eros semper dictum. Etiam tincidunt venenatis venenatis. Praesent eget gravida lorem, ut congue diam. Etiam facilisis elit at porttitor egestas. Praesent consequat, velit non vulputate convallis, ligula diam sagittis urna, in venenatis nisi justo ut mauris. Vestibulum posuere sollicitudin mi, et vulputate nisl fringilla non. Nulla ornare pretium velit a euismod. Nunc sagittis venenatis vestibulum. Nunc sodales libero a est ornare ultricies. Sed sed leo sed orci pellentesque ultrices. Mauris sollicitudin, sem quis placerat ornare, velit arcu convallis ligula, pretium finibus nisl sapien vel sem. Vivamus sit amet tortor id lorem consequat hendrerit. Nullam at dui risus.

Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia Curae; Lorem ipsum dolor sit amet, consectetur adipiscing elit. Sed feugiat semper velit consequat facilisis. Etiam facilisis justo non iaculis dictum. Fusce turpis neque, pharetra ut odio eu, hendrerit rhoncus lacus. Nunc orci felis, imperdiet vel interdum quis, porta eu ipsum. Pellentesque dictum sem lacinia, auctor dui in, malesuada nunc. Maecenas sit amet mollis eros. Proin fringilla viverra ligula, sollicitudin viverra ante sollicitudin congue. Donec mollis felis eu libero malesuada, et lacinia risus interdum.

Etiam vitae accumsan augue. Ut urna orci, malesuada ut nisi a, condimentum gravida magna. Nulla bibendum ex in vulputate sagittis. Nulla facilisi. Nullam faucibus et metus ac consequat. Quisque tempor eros velit, id mattis nibh aliquet a. Aenean tempor elit ut finibus auctor. Sed at imperdiet mauris. Vestibulum pharetra non lacus sed pulvinar. Sed pellentesque magna a eros volutpat ullamcorper. In hac habitasse platea dictumst. Donec ipsum mi, feugiat in eros sed, varius lacinia turpis. Donec vulputate tincidunt dui ac laoreet. Sed in eros dui. Pellentesque placerat tristique ligula eu finibus. Proin nec faucibus felis, eu commodo ipsum.

Integer eu hendrerit diam, sed consectetur nunc. Aliquam a sem vitae leo fermentum faucibus quis at sem. Etiam blandit, quam quis fermentum varius, ante urna ultricies lectus, vel pellentesque ligula arcu nec elit. Donec placerat ante in enim scelerisque pretium. Donec et rhoncus erat. Aenean tempor nisi vitae augue tincidunt luctus. Nam condimentum dictum ante, et laoreet neque pellentesque id. Curabitur consectetur cursus neque aliquam porta. Ut interdum nunc nec nibh vestibulum, in sagittis metus facilisis. Pellentesque feugiat condimentum metus. Etiam venenatis quam at ante rhoncus vestibulum. Maecenas suscipit congue pellentesque. Vestibulum suscipit scelerisque

JSON

- **JSON:** JSON is stand for JavaScript Object Notation. JSON is a standard text-based format for representing structured data based on JavaScript object syntax

SampleRecords.json

```
[
    {
        "trackid": "AA-1234",
        "reported_dt": "12/31/2019 23:59:59"
        "longitude": -111.12500000,
        "latitude": 33.37500000
    },
    {
        "trackid": "BB-7890",
        "reported_dt": "12/31/2019 23:59:59"
        "longitude": -113.67500000,
        "latitude": 35.87500000
    },
    {
        "trackid": "CC-4545",
        "reported_dt": "12/31/2019 23:59:59"
        "longitude": -115.57500000,
        "latitude": 37.67500000
    }
]
```

HTML

- **HTML:** HTML stands for Hyper Text Markup Language, is used for creating web pages. We can read html table in python pandas using read_html() function.

```html
<!DOCTYPE html >
<html>
<head>
<meta http-equiv="Content-Type" content="text/html; charset=utf-8" />
</head>
<body>
<div id="header">
<ul id="A">
    <li><a href="#">A link</a></li>
</ul>
<div id="B"><span class="about">About something</span></div>
</div>

<div id="main">Main container...</div>

<p id="footer">This is the footer. <a href="#">Yet another link</a></p>
</body>
</html>
```

XML

- An XML file is an **XML** (Extensible Markup Language) data file. It contains a formatted dataset that is intended to be processed by a website, web application, or software program.
- Because websites and applications can easily process XML, and humans can easily understand the data XML files contain, XML has become a standard way of transferring data over the Internet and between programs.
- Unlike **HTML**, XML allows developers to structure data using custom tags.
- This flexibility makes XML ideal for cataloging information about nearly any set of related items. For example, a developer creating a catalog of automobiles may include the following entry in their XML file:

```
<auto>
  <manufacturer>Tesla</manufacturer>
  <model>S</model>
  <horsepower>670 to 1,020</horsepower>
  <price>$69,420+</price>
</auto>
```