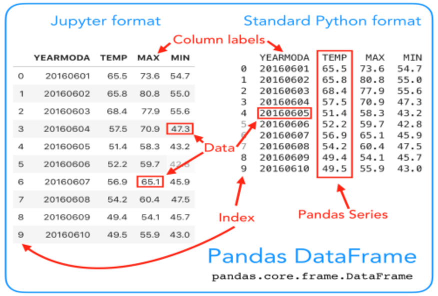
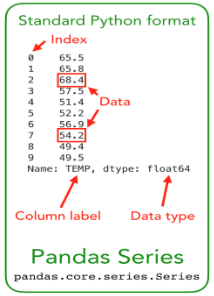
Python Pandas for Data Analysis and Manipulation



	YEARMODA	TEMP	MAX	MIN
0	20160601	65.5	73.6	54.7
1	20160602	65.8	80.8	55.0
2	20160603	68.4	77.9	55.6
3	20160604	57.5	70.9	47.3
4	20160605	51.4	58.3	43.2
5	20160606	52.2	59.7	42.8
6	20160607	56.9	65.1	45.9
7	20160608	54.2	60.4	47.5
8	20160609	49.4	54.1	45.7
9	20160610	49.5	55.9	43.0

```
   YEARMODA  TEMP  MAX   MIN
0  20160601  65.5  73.6  54.7
1  20160602  65.8  80.8  55.0
2  20160603  68.4  77.9  55.6
3  20160604  57.5  70.9  47.3
4  20160605  51.4  58.3  43.2
5  20160606  52.2  59.7  42.8
6  20160607  56.9  65.1  45.9
7  20160608  54.2  60.4  47.5
8  20160609  49.4  54.1  45.7
9  20160610  49.5  55.9  43.0
```

```
pandas.core.frame.DataFrame
```



```
0    65.5
1    65.8
2    68.4
3    57.5
4    51.4
5    52.2
6    56.9
7    54.2
8    49.4
9    49.5
Name: TEMP, dtype: float64
```

```
pandas.core.series.Series
```

Data Creation and Loading:

1. **pd.DataFrame():** Create a DataFrame from various data structures.
2. **pd.read_csv():** Read data from a CSV file.
3. **pd.read_excel():** Read data from an Excel file.
4. **pd.read_json():** Read data from a JSON file.
5. **pd.read_html():** Read data from HTML tables.

Prepared By: Syed Afroz Ali (Kaggle Grandmaster)
Follow for more AI content: https://lnkd.in/gxcsx77g

6. **pd.from_dict():** Create a DataFrame from a dictionary.
7. **pd.from_records():** Create a DataFrame from a list of records.

Data Inspection and Exploration:

1. **df.head():** View the first few rows.
2. **df.tail():** View the last few rows.
3. **df.shape():** Get the dimensions of the DataFrame.
4. **df.info():** Get information about the DataFrame.
5. **df.describe():** Get summary statistics.
6. **df.dtypes():** Get data types of columns.
7. **df.columns():** Get column names.
8. **df.index():** Get row labels.
9. **df.nunique():** Count unique values in each column.

Prepared By: Syed Afroz Ali (Kaggle Grandmaster)
Follow for more AI content: https://lnkd.in/gxcsx77g

10. **df.value_counts():** Count occurrences of unique values.
11. **df.isnull():** Check for missing values.
12. **df.notnull():** Check for non-missing values.
13. **df.corr():** Calculate correlations between columns.
14. **df.cov():** Calculate covariances between columns.
15. **df.groupby():** Group data by specified columns.
16. **df.pivot_table():** Create pivot tables.
17. **df.plot():** Create various plots.
18. **df.hist():** Create histograms.
19. **df.boxplot():** Create box plots.
20. **df.scatter_plot():** Create scatter plots.

Prepared By: Syed Afroz Ali (Kaggle Grandmaster)
Follow for more AI content: https://lnkd.in/gxcsx77g

Data Cleaning and Preparation:

1. **df.fillna():** Fill missing values.
2. **df.dropna():** Drop missing values.
3. **df.replace():** Replace values.
4. **df.astype():** Convert data types.
5. **df.rename():** Rename columns or index labels.
6. **df.drop():** Drop columns or rows.
7. **df.sample():** Sample rows randomly.
8. **df.sort_values():** Sort values.
9. **df.reset_index():** Reset the index.
10. **df.set_index():** Set a new index.

Prepared By: Syed Afroz Ali (Kaggle Grandmaster)
Follow for more AI content: https://lnkd.in/gxcsx77g

Data Manipulation and Transformation:

1. **df.loc:** Access data by labels.
2. **df.iloc:** Access data by integer locations.
3. **df.apply():** Apply a function to each element or row/column.
4. **df.map():** Map values to a dictionary or Series.
5. **df.merge():** Merge DataFrames based on common columns.
6. **df.concat():** Concatenate DataFrames along rows or columns.
7. **df.join():** Join DataFrames based on index or column.
8. **df.stack():** Convert a DataFrame from wide to long format.
9. **df.unstack():** Convert a DataFrame from long to wide format.
10. **df.melt():** Melt a DataFrame to long format.
11. **df.pivot():** Pivot a DataFrame based on specified columns.

Prepared By: Syed Afroz Ali (Kaggle Grandmaster)
Follow for more AI content: https://lnkd.in/gxcsx77g

Data Aggregation and Grouping:

1. **df.groupby().agg():** Aggregate data within groups.
2. **df.groupby().transform():** Transform data within groups.
3. **df.groupby().filter():** Filter groups based on conditions.
4. **df.groupby().apply():** Apply a function to each group.
5. **df.resample():** Resample time series data..

Data Filtering and Selection:

1. **df.query():** Filter data based on a query string.
2. **df.between():** Filter data within a range.
3. **df.isin():** Filter data based on membership in a set.
4. **df.duplicated():** Check for duplicate rows.

Prepared By: Syed Afroz Ali (Kaggle Grandmaster)
Follow for more AI content: https://lnkd.in/gxcsx77g

5. **df.drop_duplicates():** Remove duplicate rows.

Data Visualization:

1. **df.plot.bar():** Create bar plots.
2. **df.plot.line():** Create line plots.
3. **df.plot.scatter():** Create scatter plots.
4. **df.plot.pie():** Create pie charts.
5. **df.plot.hist():** Create histograms.
6. **df.plot.box():** Create box plots.
7. **df.plot.kde():** Create kernel density estimates.
8. **df.plot.hexbin():** Create hexbin plots.
9. **df.plot.density():** Create density plots.
10. **df.plot.area():** Create area plots.

Prepared By: Syed Afroz Ali (Kaggle Grandmaster)
Follow for more AI content: https://lnkd.in/gxcsx77g

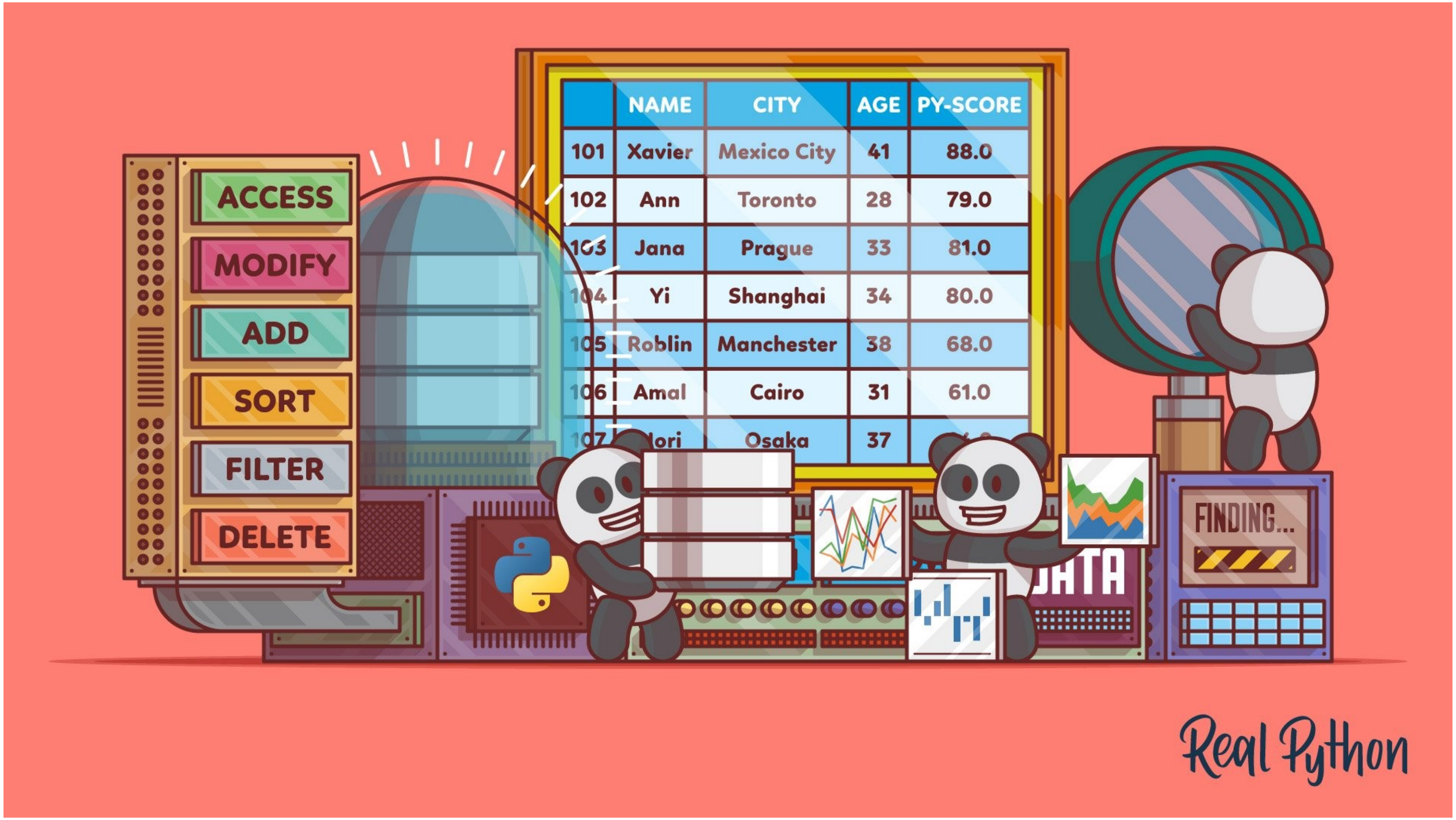
Time Series Analysis:

1. **df.shift():** Shift the index or columns.
2. **df.diff():** Calculate the difference between consecutive values.
3. **df.pct_change():** Calculate the percentage change.
4. **df.rolling():** Create rolling windows.
5. **df.expanding():** Create expanding windows.
6. **df.ewm():** Create exponentially weighted moving average windows.
7. **df.autocorr():** Calculate autocorrelation.
8. **df.corrwith():** Calculate correlations with another Series.
9. **df.detrend():** Remove trend from a time series.
10. **df.seasonal_decompose():** Decompose a time series into trend, seasonal, and residual components.

Statistical Analysis:

1. **df.skew():** Calculate skewness.
2. **df.kurtosis():** Calculate kurtosis.
3. **df.quantile():** Calculate quantiles.
4. **df.mode():** Find the most frequent values.
5. **df.rank():** Rank the values in each column.
6. **df.var():** Calculate variance.
7. **df.std():** Calculate standard deviation.
8. **df.sem():** Calculate standard error of the mean.
9. **df.median():** Calculate the median.
10. **df.mad():** Calculate the mean absolute deviation.

Prepared By: Syed Afroz Ali (Kaggle Grandmaster)
Follow for more AI content: https://lnkd.in/gxcsx77g



	NAME	CITY	AGE	PY-SCORE
101	Xavier	Mexico City	41	88.0
102	Ann	Toronto	28	79.0
103	Jana	Prague	33	81.0
104	Yi	Shanghai	34	80.0
105	Roblin	Manchester	38	68.0
106	Amal	Cairo	31	61.0
107	ori	Osaka	37	[illegible]

Prepared By: Syed Afroz Ali (Kaggle Grandmaster)
Follow for more AI content: https://lnkd.in/gxcsx77g