Pandas ETL Cheat Sheet – Complete 19 Sections (With Definitions)

1. Data Extraction (Read)

pd.read_csv(): Reads CSV file into DataFrame

pd.read_excel(): Reads Excel file into DataFrame

pd.read_json(): Reads JSON data into DataFrame

pd.read_sql(): Reads data from SQL query or table

pd.read_parquet(): Reads Parquet columnar data

pd.read_table(): Reads delimited text files

2. Initial Data Inspection

df.head(): Displays first rows

df.tail(): Displays last rows

df.sample(): Random sample of rows

df.info(): Schema, nulls, memory usage

df.describe(): Statistical summary

df.shape: Rows and columns count

df.columns: Column names

df.dtypes: Column data types

3. Column & Row Selection

df['col']: Select single column

df[['c1','c2']]: Select multiple columns

df.loc[]: Label-based selection

df.iloc[]: Index-based selection

df.at[]: Fast scalar access by label

df.iat[]: Fast scalar access by index

4. Filtering & Conditions

df.query(): SQL-like row filtering

df.isin(): Checks membership in list

df.between(): Filters range values

df.where(): Keeps values matching condition

df.mask(): Replaces values matching condition

5. Missing Data Handling

df.isna(): Detects missing values

df.notna(): Detects non-missing values

df.dropna(): Removes null rows/columns

df.fillna(): Fills missing values

df.interpolate(): Estimates missing values

df.ffill(): Forward fill

df.bfill(): Backward fill

6. Data Type Conversion

df.astype(): Converts column data types

pd.to_datetime(): Converts to datetime

pd.to_numeric(): Converts to numeric

pd.to_timedelta(): Converts to timedelta

df.convert_dtypes(): Auto-optimizes dtypes

7. Column Creation & Transformation

df.assign(): Creates new columns

df.rename(): Renames columns/index

df.apply(): Applies function row/column-wise

df.map(): Maps values in Series

df.applymap(): Applies function to each cell

8. String Operations

str.lower(): Converts to lowercase

str.upper(): Converts to uppercase

str.strip(): Trims spaces

str.replace(): Replaces substring

str.contains(): Pattern check

str.split(): Splits string

str.extract(): Regex extraction

9. Date & Time Operations

dt.year: Extracts year

dt.month: Extracts month

dt.day: Extracts day

dt.hour: Extracts hour

dt.weekday: Day of week

dt.floor(): Rounds datetime down

dt.ceil(): Rounds datetime up

10. Sorting & Ranking

df.sort_values(): Sorts by column

df.sort_index(): Sorts by index

df.rank(): Assigns rank

df.nlargest(): Top N values

df.nsmallest(): Bottom N values

11. Duplicates Handling

df.duplicated(): Finds duplicate rows

df.drop_duplicates(): Removes duplicates

12. GroupBy & Aggregation

df.groupby(): Groups data

agg(): Multiple aggregations

sum(): Sum per group

mean(): Average per group

count(): Count per group

transform(): Group-wise aligned output

filter(): Filters groups

13. Joins & Combining Data

pd.merge(): SQL-style joins

df.join(): Join using index

pd.concat(): Vertical/Horizontal concatenation

14. Reshaping Data

df.pivot(): Rows to columns

df.pivot_table(): Pivot with aggregation

df.melt(): Columns to rows

df.stack(): Columns to index

df.unstack(): Index to columns

15. Measures of Dispersion

df.var(): Variance

df.std(): Standard deviation

df.mad(): Mean absolute deviation

df.quantile(): Percentiles

df.skew(): Skewness

df.kurt(): Kurtosis

df.max() - df.min(): Range

16. Rolling & Window Functions

df.rolling(): Rolling window calculations

df.expanding(): Expanding window calculations

df.cumsum(): Cumulative sum

df.cumprod(): Cumulative product

17. Conditional Logic

np.where(): Vectorized if-else

np.select(): Multiple conditions

df.where(): Conditional retain

df.mask(): Conditional replace

18. Data Validation & Quality Checks

df.nunique(): Unique value count

df.value_counts(): Frequency distribution

df.is_unique: Checks uniqueness

df.equals(): Compares DataFrames

df.compare(): Shows differences

19. Data Loading (Export)

df.to_csv(): Writes to CSV

df.to_excel(): Writes to Excel

df.to_json(): Writes to JSON

df.to_sql(): Writes to database

df.to_parquet(): Writes to Parquet