NumPy & Pandas Revision Cheat Sheet

NumPy Quick Revision

Array Creation:
- np.array([1,2,3]) → Create array
- np.zeros((m,n)), np.ones((m,n)), np.eye(n)
- np.arange(start, stop, step), np.linspace(start, stop, num)
- np.random.rand(m,n), np.random.randn(m,n), np.random.randint(low, high, size)

Array Properties:
- arr.shape, arr.ndim, arr.size, arr.dtype

Operations:
- np.add(a,b), np.subtract(a,b), np.multiply(a,b), np.divide(a,b)
- np.dot(a,b), np.cross(a,b)
- np.sqrt(arr), np.power(arr,n), np.exp(arr), np.log(arr)

Aggregation:
- np.sum(arr), np.mean(arr), np.median(arr), np.std(arr), np.var(arr)
- np.min(arr), np.max(arr), np.argmin(arr), np.argmax(arr)

Reshaping:
- arr.reshape(m,n), arr.flatten(), np.ravel(arr)
- arr.T, np.concatenate([a,b], axis=0/1)

Indexing:
- arr[:,0], arr[0,:], arr[arr>5], np.where(arr>5)

Linear Algebra:
- np.linalg.inv(a), np.linalg.det(a), np.linalg.eig(a)
- np.linalg.norm(a), np.linalg.solve(A,b)

Pandas Quick Revision

Data Structures:
- Series: 1D labeled array
- DataFrame: 2D labeled table

Input/Output:
- pd.read_csv(), pd.read_excel(), pd.read_sql(), pd.read_json()
- df.to_csv(), df.to_excel(), df.to_sql()

Inspect:
- df.head(), df.tail(), df.shape, df.info(), df.describe()
- df.dtypes, df.memory_usage()

Selection:
- df['col'], df[['c1','c2']]
- df.loc[0], df.iloc[0]
- df[df['Age']>25], df.loc[df['Age']>25, ['Name','Age']]

Cleaning:
- df.dropna(), df.fillna(val)
- df.duplicated(), df.drop_duplicates()
- df.replace(), df.rename(), df.astype()
- df.apply(func), df.map(func)

Sorting & Grouping:
- df.sort_values('col'), df.sort_index()
- df.groupby('col').mean(), df.groupby(['c1','c2']).agg({'x':'sum'})

Merging:
- pd.concat([df1,df2], axis=0/1)
- pd.merge(df1,df2,on='id', how='left/right/inner/outer')
- df1.join(df2)

Reshape:
- df.pivot(), df.pivot_table()
- df.melt(), df.stack(), df.unstack()

Time Series:
- pd.to_datetime(df['date'])
- df['date'].dt.year / month
- df.set_index('date').resample('M').sum()

Window Functions:
- df['col'].rolling(3).mean()
- df['col'].expanding().sum()
- df['col'].ewm(span=3).mean()

Other:
- df.value_counts(), df.nunique()
- df.corr(), df.cov()
- df.sample(n=5), df.isin([1,2,3])
- df.cumsum(), df.cumprod()