Pandas Cheatsheet

Importing & Basics

```
import pandas as pd
df = pd.read_csv("filename.csv")
df = pd.read_excel("filename.xlsx", sheet_name='Sheet1')
data = {'col1': [1, 2], 'col2': [3, 4]}
df = pd.DataFrame(data)
df.to_csv("filename.csv", index=False)
df.head()  # First 5 rows
df.tail(3)  # Last 3 rows
df.shape
df.info()
df.describe()
```

Selection & Filtering

```
df['col1']
df[['col1', 'col2']]
df.iloc[0]
df.iloc[0:5]
df.loc[0, 'col1']
df[df['col1'] > 10]
df[(df['col1'] > 10) & (df['col2'] == 'A')]
```

Modifying Data

```
df['new_col'] = df['col1'] * 2
df.rename(columns={'old': 'new'}, inplace=True)
df['col1'].replace({1: 'A', 2: 'B'}, inplace=True)
df.drop(columns=['col1'], inplace=True)
df.drop(index=0, inplace=True)
df['col1'] = df['col1'].astype('float')
df['col1'] = df['col1'].apply(lambda x: x + 1)
```

Missing Data

```
df.isnull().sum()
df.dropna(inplace=True)
df.fillna(0, inplace=True)
df.fillna(method='ffill')
```

Grouping & Aggregation

```
df.groupby('col1').mean()
df.groupby(['col1', 'col2'])['col3'].sum()
df['col1'].value_counts()
```

Pandas Cheatsheet

Sorting & Duplicates

```
df.sort_values(by='col1', ascending=False)
df.drop_duplicates()
```

Merging & Joining

```
pd.merge(df1, df2, on='key')
df1.join(df2.set_index('key'), on='key')
pd.concat([df1, df2], axis=0)
pd.concat([df1, df2], axis=1)
```

Datetime Operations

```
df['date'] = pd.to_datetime(df['date'])
df['year'] = df['date'].dt.year
df['month'] = df['date'].dt.month
```

Pivot Tables & Crosstab

```
df.pivot_table(index='col1', columns='col2', values='col3', aggfunc='sum')
pd.crosstab(df['col1'], df['col2'])
```

Useful Utilities

```
df.columns
df.index
df.memory_usage()
df.sample(5)
df.nunique()
```