Python Data Analyst Cheat Sheet

BASIC PYTHON CONCEPTS

```
# Loops
for i in range(5):
    print(i)

# If-else
if x > 10:
    print("High")
else:
    print("Low")

# Function
def greet(name):
    return f"Hello, {name}"
```

PANDAS DATA WRANGLING

```
import pandas as pd

# Read file
df = pd.read_csv('data.csv')

# Quick look
df.head()
df.info()

# Filter rows
df[df['column'] == 'value']

# Select columns
df[['col1', 'col2']]

# Create new column
df['new_col'] = df['old_col'] * 10

# Group by
df.groupby('group_col')['value_col'].mean()

# Merge
merged = pd.merge(df1, df2, on='id', how='inner')

# Handle missing
df.dropna()  # or df.fillna(0)
```

NUMPY BASICS

Python Data Analyst Cheat Sheet

```
import numpy as np

arr = np.array([1, 2, 3])
arr.mean()
arr.sum()
```

BASIC VISUALS (Matplotlib & Seaborn)

```
import matplotlib.pyplot as plt
import seaborn as sns

# Line plot
plt.plot([1, 2, 3], [4, 5, 6])
plt.show()

# Histogram
sns.histplot(df['column'])

# Bar plot
sns.barplot(x='category', y='value', data=df)
```

SQL IN PYTHON

```
import sqlite3
conn = sqlite3.connect('my_db.db')
query = "SELECT * FROM table_name WHERE column > 10"
df = pd.read_sql(query, conn)
```

AB TESTING SNIPPETS

```
from statsmodels.stats.proportion import proportions_ztest

successes = [success_ctrl, success_treat]
nobs = [n_ctrl, n_treat]

z_stat, pval = proportions_ztest(successes, nobs=nobs)
```

STATISTICS BASICS

```
from scipy import stats

# T-test
stats.ttest_ind(group1, group2)
```

Python Data Analyst Cheat Sheet

```
# Chi-square
stats.chisquare([obs1, obs2], f_exp=[exp1, exp2])
```

AUTOMATION / LOOPS

```
# Apply function to column
df['new_col'] = df['col'].apply(lambda x: x * 2)

# For loop over columns
for col in df.columns:
    print(df[col].mean())
```

CLEANING + STRINGS

```
# Lowercase column names
df.columns = df.columns.str.lower()

# Strip whitespaces
df['column'] = df['column'].str.strip()
```

TO REMEMBER

```
Task                      | Tool           | Example
--------------------------|----------------|------------------------------
Data analysis             | pandas         | df.groupby(...)
Visualization             | seaborn/matplt | sns.barplot(...)
Statistical testing       | statsmodels    | proportions_ztest(...)
SQL integration           | sqlite         | pd.read_sql(...)
Export to Excel           | pandas         | df.to_excel('file.xlsx')
```