DSC 80 Discussion 1 Worksheet

Name:

1 WI24 Midterm Problem 1

	owner_id	owner_age	owner_sex	district	primary_breed	secondary_breed	dog_sex	birth_year
0	4215	41-50	f	8	Bergamasker	NaN	f	2004
1	4215	41-50	f	8	Border Collie	NaN	m	2001
2	6071	61-70	m	3	Cocker Spaniel	Labrador Retriever	m	2014
3	123237	21-30	f	7	Sheltie	NaN	m	2014
4	135726	11-20	f	11	Pinscher	NaN	f	2016

Write an expression that evaluates to a Series containing the primary breeds of all female dogs.

```
dogs.loc[dogs['dog_sex'] == 'f', 'primary_breed']
```

Write an expression that evaluates to the most common district in `dogs`. Assume there are no ties.

```
dogs['district'].value_counts().index[0]    OR    dogs['district'].value_counts().idxmax()
```

2 FA22 Midterm Problem 3

	category	completed	minutes	urgency	client
0	work	False	NaN	2.0	NaN
1	work	False	NaN	1.0	NaN
2	work	True	13.5	2.0	NaN
3	work	False	NaN	1.0	NaN
4	relationship	True	5.3	NaN	NaN

Define a `personal` task to be one whose category is not `work` or `consulting`. Write a piece of code which adds a new column named `personal` to tasks. An entry of the column should be True if the task is a `personal` task, and False otherwise.

```
tasks['personal'] = (tasks['category'] != 'work') & (tasks['category'] == 'consulting')

OR

tasks['personal'] = ~tasks['category'].isin(['work', 'consulting'])
```

3 NBA Data

Player	Tm	Pos	Age	G	PTS	AST	TRB	FG%
Joel Embiid	PHI	C	28	66	33.1	4.2	10.2	0.548
Luka Doncic	DAL	PG	23	66	32.4	8.0	8.6	0.496
Damian Lillard	POR	PG	32	58	32.2	7.3	4.8	0.463
Shai Gilgeous-Alexander	OKC	PG	24	68	31.4	5.5	4.8	0.510
Giannis Antetokounmpo	MIL	PF	28	63	31.1	5.7	11.8	0.553

(more rows follow)

The above data is from the 2022-23 NBA season, indexed by player names, with columns for their team, position, age, games played, and averages of points, assists, rebounds, and field goal percentage.

For the following questions, assume the DataFrame is stored in the variable `nba`. Note that the dataset has been sorted by descending points per game (PTS), but the indices have not been changed.

Write a line of code that returns (as an `int`) the number of games that Damian Lillard played:

```
nba.loc['Damian Lillard']['G']
```

Write a line that returns a DataFrame of games played and points average for just the top three scorers:

```
nba.iloc[:3][['G', 'PTS']]
```

Write a line that returns a subset of `nba` of players under the age of 22 who score at least 20 points per game:

```
nba[(nba['age'] < 22) & (nba['PTS'] >= 20)]
```

Write a line that adds a new column `box_sum` to `nba` containing the sum of `PTS, AST,` and `REB`, *without* using a `+` sign:

```
nba['box_sum'] = nba[['PTS', 'AST', 'TRB']].sum(axis = 1)
```

Write a line that returns the name of the lowest-scoring player on the Milwaukee Bucks (team `MIL`). Remember, the data is sorted by *descending* scoring averages!

```
nba[nba['Tm'] == 'MIL']].index[-1]
```