Import pandas to access and read the CSV file.

```python
import pandas as pd
coffee = pd.read_csv("index_1.csv")
coffee.head(10)
```

	date	datetime	cash_type	card	money	coffee_name
0	2024-03-01	2024-03-01 10:15:50.520	card	ANON-0000-0000-0001	38.7	Latte
1	2024-03-01	2024-03-01 12:19:22.539	card	ANON-0000-0000-0002	38.7	Hot Chocolate
2	2024-03-01	2024-03-01 12:20:18.089	card	ANON-0000-0000-0002	38.7	Hot Chocolate
3	2024-03-01	2024-03-01 13:46:33.006	card	ANON-0000-0000-0003	28.9	Americano
4	2024-03-01	2024-03-01 13:48:14.626	card	ANON-0000-0000-0004	38.7	Latte
5	2024-03-01	2024-03-01 15:39:47.726	card	ANON-0000-0000-0005	33.8	Americano with Milk
6	2024-03-01	2024-03-01 16:19:02.756	card	ANON-0000-0000-0006	38.7	Hot Chocolate
7	2024-03-01	2024-03-01 18:39:03.580	card	ANON-0000-0000-0007	33.8	Americano with Milk
8	2024-03-01	2024-03-01 19:22:01.762	card	ANON-0000-0000-0008	38.7	Cocoa
9	2024-03-01	2024-03-01 19:23:15.887	card	ANON-0000-0000-0008	33.8	Americano with Milk

Next steps: Generate code with `coffee`   New interactive sheet

Count how many rows and columns in the dataset

```python
coffee.shape
```

```
(3636, 6)
```

Check dataset details and duplicates

```python
coffee.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3636 entries, 0 to 3635
Data columns (total 6 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   date         3636 non-null   object
 1   datetime     3636 non-null   object
 2   cash_type    3636 non-null   object
 3   card         3547 non-null   object
 4   money        3636 non-null   float64
 5   coffee_name  3636 non-null   object
dtypes: float64(1), object(5)
memory usage: 170.6+ KB
```

```python
coffee.duplicated().sum()
```

```
np.int64(0)
```

Check and Handle Missing Values

```python
coffee.isnull().any()
```

	0
date	False
datetime	False
cash_type	False
card	True
money	False
coffee_name	False

dtype: bool

```
coffee["card"] = coffee["card"].fillna("card_payment")
coffee.head()
```

	date	datetime	cash_type	card	money	coffee_name	⊞
0	2024-03-01	2024-03-01 10:15:50.520	card	ANON-0000-0000-0001	38.7	Latte	
1	2024-03-01	2024-03-01 12:19:22.539	card	ANON-0000-0000-0002	38.7	Hot Chocolate	
2	2024-03-01	2024-03-01 12:20:18.089	card	ANON-0000-0000-0002	38.7	Hot Chocolate	
3	2024-03-01	2024-03-01 13:46:33.006	card	ANON-0000-0000-0003	28.9	Americano	
4	2024-03-01	2024-03-01 13:48:14.626	card	ANON-0000-0000-0004	38.7	Latte	

Next steps: Generate code with `coffee`   New interactive sheet

```
coffee.isna().any()
```

	0
date	False
datetime	False
cash_type	False
card	False
money	False
coffee_name	False

dtype: bool

Start coding or generate with AI.

˅ Fix Data Types for Date and Time Columns

```
coffee.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3636 entries, 0 to 3635
Data columns (total 6 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   date         3636 non-null   object
 1   datetime     3636 non-null   object
 2   cash_type    3636 non-null   object
 3   card         3636 non-null   object
 4   money        3636 non-null   float64
 5   coffee_name  3636 non-null   object
dtypes: float64(1), object(5)
memory usage: 170.6+ KB
```

```
coffee["date"] = pd.to_datetime(coffee["date"])
coffee["datetime"] = pd.to_datetime(coffee["datetime"])
coffee.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3636 entries, 0 to 3635
Data columns (total 6 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   date         3636 non-null   datetime64[ns]
 1   datetime     3636 non-null   datetime64[ns]
```

```
 2   cash_type    3636 non-null   object
 3   card         3636 non-null   object
 4   money        3636 non-null   float64
 5   coffee_name  3636 non-null   object
dtypes: datetime64[ns](2), float64(1), object(3)
memory usage: 170.6+ KB
```

Feature Engineering from Date & Time

```python
coffee["hour_of_day"] = coffee["datetime"].dt.hour
coffee["day_of_week"] = coffee["date"].dt.day_name()
coffee["is_weekend"] = coffee["day_of_week"].isin(["saturday", "sunday"])
coffee.head()
```

	date	datetime	cash_type	card	money	coffee_name	hour_of_day	day_of_week	is_weekend
0	2024-03-01	2024-03-01 10:15:50.520	card	ANON-0000-0000-0001	38.7	Latte	10	Friday	False
1	2024-03-01	2024-03-01 12:19:22.539	card	ANON-0000-0000-0002	38.7	Hot Chocolate	12	Friday	False
2	2024-03-01	2024-03-01 12:20:18.089	card	ANON-0000-0000-0002	38.7	Hot Chocolate	12	Friday	False
3	2024-03-01	2024-03-01 13:46:33.006	card	ANON-0000-0000-0003	28.9	Americano	13	Friday	False
4	2024-03-01	2024-03-01 13:48:14.626	card	ANON-0000-0000-0004	38.7	Latte	13	Friday	False

Next steps: (Generate code with `coffee`) (New interactive sheet)

```python
coffee.rename(columns={"cash_type": "payment_method"}, inplace=True)
coffee.head()
```

	date	datetime	card	money	coffee_name	hour_of_day	day_of_week	is_weekend	payment_method
0	2024-03-01	2024-03-01 10:15:50.520	ANON-0000-0000-0001	38.7	Latte	10	Friday	False	card
1	2024-03-01	2024-03-01 12:19:22.539	ANON-0000-0000-0002	38.7	Hot Chocolate	12	Friday	False	card
2	2024-03-01	2024-03-01 12:20:18.089	ANON-0000-0000-0002	38.7	Hot Chocolate	12	Friday	False	card
3	2024-03-01	2024-03-01 13:46:33.006	ANON-0000-0000-0003	28.9	Americano	13	Friday	False	card
4	2024-03-01	2024-03-01 13:48:14.626	ANON-0000-0000-0004	38.7	Latte	13	Friday	False	card

Next steps: (Generate code with `coffee`) (New interactive sheet)

```python
coffee = coffee.drop(columns=["card"])
coffee.head()
```

	date	datetime	money	coffee_name	hour_of_day	day_of_week	is_weekend	payment_method
0	2024-03-01	2024-03-01 10:15:50.520	38.7	Latte	10	Friday	False	card
1	2024-03-01	2024-03-01 12:19:22.539	38.7	Hot Chocolate	12	Friday	False	card
2	2024-03-01	2024-03-01 12:20:18.089	38.7	Hot Chocolate	12	Friday	False	card
3	2024-03-01	2024-03-01 13:46:33.006	28.9	Americano	13	Friday	False	card
4	2024-03-01	2024-03-01 13:48:14.626	38.7	Latte	13	Friday	False	card

Next steps: (Generate code with `coffee`) (New interactive sheet)

Standardise Categorical Values

```python
coffee = coffee[
    [
        "date",
        "datetime",
```

```
        "hour_of_day",
        "day_of_week",
        "is_weekend",
        "coffee_name",
        "payment_method",
        "money",
    ]
]
coffee.head()
```

	date	datetime	hour_of_day	day_of_week	is_weekend	coffee_name	payment_method	money
0	2024-03-01	2024-03-01 10:15:50.520	10	Friday	False	Latte	card	38.7
1	2024-03-01	2024-03-01 12:19:22.539	12	Friday	False	Hot Chocolate	card	38.7
2	2024-03-01	2024-03-01 12:20:18.089	12	Friday	False	Hot Chocolate	card	38.7
3	2024-03-01	2024-03-01 13:46:33.006	13	Friday	False	Americano	card	28.9
4	2024-03-01	2024-03-01 13:48:14.626	13	Friday	False	Latte	card	38.7

Next steps: Generate code with `coffee`   New interactive sheet

```
coffee["coffee_name"] = coffee["coffee_name"].str.title().str.strip()
coffee.head()
```

	date	datetime	hour_of_day	day_of_week	is_weekend	coffee_name	payment_method	money
0	2024-03-01	2024-03-01 10:15:50.520	10	Friday	False	Latte	card	38.7
1	2024-03-01	2024-03-01 12:19:22.539	12	Friday	False	Hot Chocolate	card	38.7
2	2024-03-01	2024-03-01 12:20:18.089	12	Friday	False	Hot Chocolate	card	38.7
3	2024-03-01	2024-03-01 13:46:33.006	13	Friday	False	Americano	card	28.9
4	2024-03-01	2024-03-01 13:48:14.626	13	Friday	False	Latte	card	38.7

Next steps: Generate code with `coffee`   New interactive sheet

```
coffee = coffee.drop(columns=["datetime"])
coffee.head()
```

	date	hour_of_day	day_of_week	is_weekend	coffee_name	payment_method	money
0	2024-03-01	10	Friday	False	Latte	card	38.7
1	2024-03-01	12	Friday	False	Hot Chocolate	card	38.7
2	2024-03-01	12	Friday	False	Hot Chocolate	card	38.7
3	2024-03-01	13	Friday	False	Americano	card	28.9
4	2024-03-01	13	Friday	False	Latte	card	38.7

Next steps: Generate code with `coffee`   New interactive sheet