

```
In [3]: import pandas as pd
import matplotlib.pyplot as plt
```

```
In [10]: df = pd.read_excel('winter_data_challenge.xlsx')
print (df)
```

	order_id	shop_id	user_id	order_amount	total_items	payment_method	\
0	1	53	746	224	2	cash	
1	2	92	925	90	1	cash	
2	3	44	861	144	1	cash	
3	4	18	935	156	1	credit_card	
4	5	18	883	156	1	credit_card	
...	
4995	4996	73	993	330	2	debit	
4996	4997	48	789	234	2	cash	
4997	4998	56	867	351	3	cash	
4998	4999	60	825	354	2	credit_card	
4999	5000	44	734	288	2	debit	

	created_at
0	2017-03-13 12:36:56.190
1	2017-03-03 17:38:51.999
2	2017-03-14 04:23:55.595
3	2017-03-26 12:43:36.649
4	2017-03-01 04:35:10.773
...	...
4995	2017-03-30 13:47:16.597
4996	2017-03-16 20:36:16.389
4997	2017-03-19 05:42:42.228
4998	2017-03-16 14:51:18.188
4999	2017-03-18 15:48:18.205

[5000 rows x 7 columns]

```
In [11]: df.head() #checking the first few rows
```

	order_id	shop_id	user_id	order_amount	total_items	payment_method	created_at
0	1	53	746	224	2	cash	2017-03-13 12:36:56.190
1	2	92	925	90	1	cash	2017-03-03 17:38:51.999
2	3	44	861	144	1	cash	2017-03-14 04:23:55.595
3	4	18	935	156	1	credit_card	2017-03-26 12:43:36.649
4	5	18	883	156	1	credit_card	2017-03-01 04:35:10.773

```
In [13]: df.isnull().sum() #checking if there is any null value
```

```
Out[13]: order_id      0
shop_id      0
user_id      0
order_amount  0
total_items  0
payment_method 0
created_at    0
dtype: int64
```

```
In [14]: df.duplicated() #checking duplicates
```

```
Out[14]: 0      False
```