

# Diego Lozoya Morales | 745345

If a step is a question, answer the question after you solve it! ->  
`print("This is the answer")`

```
In [1]: import pandas as pd
```

Import all the files

```
In [2]: visits = pd.read_csv('Data/visits-1.csv',  
                             parse_dates=[1])  
  
cart = pd.read_csv('Data/cart-1.csv',  
                   parse_dates=[1])  
  
checkout = pd.read_csv('Data/checkout-1.csv',  
                        parse_dates=[1])  
  
purchase = pd.read_csv('Data/purchase-1.csv',  
                        parse_dates=[1])
```

Step 1: Inspect the DataFrames using `print` and `head`

```
In [3]: print(visits.head())  
print(cart.head())  
print(checkout.head())  
print(purchase.head())
```

	user_id	visit_time
0	943647ef-3682-4750-a2e1-918ba6f16188	2017-04-07 15:14:00
1	0c3a3dd0-fb64-4eac-bf84-ba069ce409f2	2017-01-26 14:24:00
2	6e0b2d60-4027-4d9a-babd-0e7d40859fb1	2017-08-20 08:23:00
3	6879527e-c5a6-4d14-b2da-50b85212b0ab	2017-11-04 18:15:00
4	a84327ff-5daa-4ba1-b789-d5b4caf81e96	2017-02-27 11:25:00

  

	user_id	cart_time
0	2be90e7c-9cca-44e0-bcc5-124b945ff168	2017-11-07 20:45:00
1	4397f73f-1da3-4ab3-91af-762792e25973	2017-05-27 01:35:00
2	a9db3d4b-0a0a-4398-a55a-ebb2c7adf663	2017-03-04 10:38:00
3	b594862a-36c5-47d5-b818-6e9512b939b3	2017-09-27 08:22:00
4	a68a16e2-94f0-4ce8-8ce3-784af0bbb974	2017-07-26 15:48:00

  

	user_id	checkout_time
0	d33bdc47-4afa-45bc-b4e4-dbe948e34c0d	2017-06-25 09:29:00
1	4ac186f0-9954-4fea-8a27-c081e428e34e	2017-04-07 20:11:00
2	3c9c78a7-124a-4b77-8d2e-e1926e011e7d	2017-07-13 11:38:00
3	89fe330a-8966-4756-8f7c-3bdbcd47279a	2017-04-20 16:15:00
4	3ccdaf69-2d30-40de-b083-51372881aedd	2017-01-08 20:52:00

  

	user_id	purchase_time
0	4b44ace4-2721-47a0-b24b-15fbfa2abf85	2017-05-11 04:25:00
1	02e684ae-a448-408f-a9ff-dcb4a5c99aac	2017-09-05 08:45:00
2	4b4bc391-749e-4b90-ab8f-4f6e3c84d6dc	2017-11-20 20:49:00
3	a5dbb25f-3c36-4103-9030-9f7c6241cd8d	2017-01-22 15:18:00
4	46a3186d-7f5a-4ab9-87af-84d05bfd4867	2017-06-11 11:32:00

Step 2: Left merging visits and cart

```
In [4]: visits_cart = visits.merge(cart, how='left')
```

Step 3: How long is `visits_cart` ?

```
In [5]: print(len(visits_cart))
```

2000

Step 4: How many timestamps are null for `cart_time` ?

```
In [6]: null_cart_time = visits_cart.cart_time.isnull().sum()
print(null_cart_time)
```

1652

Step 5: What percentage only visited?

```
In [7]: visited_not_cart = null_cart_time / len(visits_cart)
print(f'{visited_not_cart:.2%}')
```

82.60%

Step 6: What percentage placed a t-shirt in their cart but did not checkout?

```
In [8]: cart.merge(checkout, how='left').head()
```

Out [8]:

	user_id	cart_time	checkout_time
0	2be90e7c-9cca-44e0-bcc5-124b945ff168	2017-11-07 20:45:00	2017-11-07 21:14:00
1	4397f73f-1da3-4ab3-91af-762792e25973	2017-05-27 01:35:00	NaT
2	a9db3d4b-0a0a-4398-a55a-ebb2c7adf663	2017-03-04 10:38:00	2017-03-04 11:04:00
3	b594862a-36c5-47d5-b818-6e9512b939b3	2017-09-27 08:22:00	2017-09-27 08:26:00
4	a68a16e2-94f0-4ce8-8ce3-784af0bbb974	2017-07-26 15:48:00	NaT

```
In [9]: cart_checkout = cart.merge(checkout, how='left')

no_checkout_time = cart_checkout.checkout_time.isnull().sum()
no_checkout_time_pct = no_checkout_time / len(cart_checkout)

print(f"{no_checkout_time_pct:.2%} of shoppers placed a t-shirt in their car
```

35.06% of shoppers placed a t-shirt in their cart but did not checkout.

Step 7: Merge it all together

```
In [10]: all_data = visits.merge(cart, how='left').merge(checkout, how='left').merge(
```

Step 8: % of users who got to checkout but did not purchase

```
In [11]: checkout_users = all_data[all_data.checkout_time.notna()]
checkout_not_purchase = checkout_users.purchase_time.isna().sum()
checkout_not_purchase_pct = checkout_not_purchase / len(checkout_users)

print(f"{checkout_not_purchase_pct:.2%} of shoppers checked out but did not
```

24.55% of shoppers checked out but did not purchase a t-shirt.

Step 9: check each part of the funnel, let's print all 3 of them again

```
In [12]: print("{} percent of users who visited the page did not add a t-shirt to the
print("{} percent of users who added a t-shirt to their cart did not checkou
print("{} percent of users who made it to checkout did not purchase a shirt
```

82.6 percent of users who visited the page did not add a t-shirt to their cart

35.06 percent of users who added a t-shirt to their cart did not checkout

24.55 percent of users who made it to checkout did not purchase a shirt

*The weakest part of the funnel is clearly getting a person who visited the site to add a tshirt to their cart. Once they've added a t-shirt to their cart it is fairly likely they end up purchasing it. A suggestion could be to make the add-to-cart button more prominent on the front page.*

Step 10: adding new column

```
In [13]: all_data['time_to_purchase'] = all_data.purchase_time - all_data.visit_time
```

Step 11: examine the results

```
In [14]: print(all_data.time_to_purchase.dropna())

1      0 days 00:44:00
14     0 days 00:38:00
48     0 days 00:41:00
49     0 days 01:00:00
65     0 days 01:00:00

...
2083   0 days 00:29:00
2093   0 days 00:46:00
2097   0 days 00:32:00
2098   0 days 00:28:00
2099   0 days 00:11:00
Name: time_to_purchase, Length: 252, dtype: timedelta64[ns]
```

Step 12: what is the average time to purchase?

```
In [15]: print(all_data.time_to_purchase.mean())

0 days 00:43:12.380952380
```