

b. Create a small df sample

Creating a sample of df `opc_merge` because keep getting "not enough memory" error messages after running the for-loop statements

```
# assigning random values to create a 70/30 sample
```

```
np.random.seed(4)
dev = np.random.rand(len(opc_merge_keep)) <= 0.7
```

```
# splitting into two sample groups, small and big
```

```
big = opc_merge_keep[dev]
small = opc_merge_keep[~dev]
```

```
len(dev)
```

30199150

```
len(small)
```

9058349

```
len(big)
```

21140801

```
len(small)+len(big)
```

30199150

```
len(opc_merge_keep)
```

30199150

```
# make new df
```

```
opc_merge_keep_sample = small
```

```
# checking all columns are still there
```

```
opc_merge_keep_sample.head(1)
```

	order_id	user_id	order_number	order_day_of_week	order_hour_of_day	days_since_prior_order	product_id	add_to_cart_order
0	2539329	1	1	2	8	NaN	196	

1 rows × 36 columns

```
# checking if number of rows match above count
```

```
opc_merge_keep_sample.shape
```

(9058349, 36)

```
# exporting subset sample df 'opc_merge_keep_sample'
```

```
opc_merge_keep_sample.to_pickle(os.path.join(path, '02 Data', 'Prepared Data', 'orders_products_customer
```