› Establishing Python Library Packages

Show code

› Dataset Overview

Show code

	InvoiceNo	StockCode	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	Descript
0	536365	85123A	6	2010-12-01 08:26:00	2.55	17850.0	United Kingdom	WH HANG HEAR LI HOL
1	536365	71053	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom	WH ME

› Add Column: Date

Show code

	InvoiceNo	StockCode	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	Descript
0	536365	85123A	6	2010-12-01 08:26:00	2.55	17850.0	United Kingdom	WH HANG HEAR LI HOL
1	536365	71053	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom	WH ME LANT
2	536365	84406B	8	2010-12-01 08:26:00	2.75	17850.0	United Kingdom	CR CU HEA

› Check Date Range

Show code

	Quantity	InvoiceDate	UnitPrice	CustomerID
count	533678.000000	533678	533678.000000	399764.000000
mean	9.686862	2011-07-04 10:48:19.647577344	3.289060	15288.638562
min	-80995.000000	2010-12-01 08:26:00	0.000000	12346.000000
25%	1.000000	2011-03-28 11:34:00	1.250000	13959.000000
50%	3.000000	2011-07-19 15:23:00	2.080000	15152.000000
75%	10.000000	2011-10-18 17:10:00	4.130000	16791.000000
max	80995.000000	2011-12-09 12:50:00	1867.860000	18287.000000
std	219.446826	NaN	5.280527	1710.727958

DATE RANGE OF RECORDED TRANSACTIONS

- Date of First Transaction on Record: 2010 Dec 01
- Date of Last Transaction on Record: 2011 Dec 09

> Add Column: Revenue per row

Show code

	InvoiceNo	StockCode	Quantity	InvoiceDate	UnitPrice	CustomerID	Country	Descript
0	536365	85123A	6	2010-12-01 08:26:00	2.55	17850.0	United Kingdom	WH HANG HEAR LI HOL
1	536365	71053	6	2010-12-01 08:26:00	3.39	17850.0	United Kingdom	WH ME LANT
2	536365	84406B	8	2010-12-01 08:26:00	2.75	17850.0	United Kingdom	CR CU HEA

⌄ Customer Segmentation: RFM Modeling

- Recency: how recently a customer made their most recent purchase (measure the time elapsed)

- Frequency: number of times a customer has made a transaction within a certain time frame

  > in this dataset, use unique 'InvoiceNo' values, as each represents the distinct transactions, even if multiple items are included in the same invoice

- Monetary: total amount of money a customer has spent over a specific period

> ## Generate RFM DataFrame

Show code

CustomerID	Recency	Frequency	Monetary
18102.0	1	60	259657.30
14422.0	1	6	4263.64
12680.0	1	4	790.81
12748.0	1	217	29819.99
17001.0	1	11	3989.57
...	...	...	...
14142.0	374	1	311.81
17908.0	374	1	232.03
17968.0	374	1	265.10
12791.0	374	1	177.60
16583.0	374	1	233.45

4363 rows × 3 columns

∨ Customer Ranking

ranking customers per RFM attribute
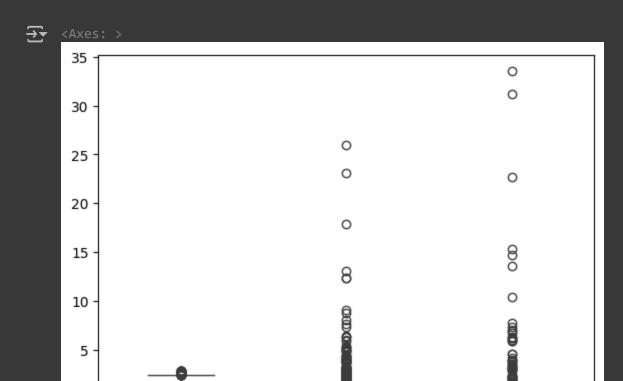
> ## Add Columns: customer rank per RFM attribute

Show code

Show code

<Axes: >