

# DATA ANALYST INTERNSHIP

## Task 1: Data Cleaning and Preprocessing

### Selected Dataset: Netflix Movies and TV Shows

Raw Data:

Row ID	Order ID	Order Date	Ship Date	Ship Mode	Customer	Customer Segment	Country	City	State	Postal Code	Region	Product ID	Category	Sub-Category	Product Name	Sales	Quantity	Discount	Profit
1	CA-2013-1	9/11/2013	12/11/2013	Second Class	CG-12520	Claire Gut Consumer	United States	Henderson	Kentucky	42420	South	FUR-BO-1C	Furniture	Bookcases	Bush Somerset	261.96	2	0	41.9136
2	CA-2013-1	9/11/2013	12/11/2013	Second Class	CG-12520	Claire Gut Consumer	United States	Henderson	Kentucky	42420	South	FUR-CH-1C	Furniture	Chairs	Hon Delux	731.94	3	0	219.582
3	CA-2013-113-06-2013	17-06-2013		Second Class	DV-13045	Darrin Vai Corporate	United States	Los Angeles	California	90036	West	OFF-IA-10	Office Supplies	Labels	Self-Adhes	14.62	2	0	6.8714
4	US-2012-1	11/10/2012	18-10-2012	Standard	SO-20335	Sean O'Do Consumer	United States	Fort Lauderdale	Florida	33311	South	FUR-TA-10	Furniture	Tables	Bretford C	957.578	5	0.45	-383.031
5	US-2012-1	11/10/2012	18-10-2012	Standard	SO-20335	Sean O'Do Consumer	United States	Fort Lauderdale	Florida	33311	South	OFF-ST-10	Office Supplies	Storage	Eldon Folc	22.368	2	0.2	2.5164
6	CA-2011-1	9/6/2011	14-06-2011	Standard	BH-11710	Brosina H Consumer	United States	Los Angeles	California	90032	West	FUR-FU-1C	Furniture	Furnishings	Eldon Expi	48.86	7	0	14.1694
7	CA-2011-1	9/6/2011	14-06-2011	Standard	BH-11710	Brosina H Consumer	United States	Los Angeles	California	90032	West	OFF-AR-1C	Office Supplies	Art	Newell 32	7.28	4	0	1.9656
8	CA-2011-1	9/6/2011	14-06-2011	Standard	BH-11710	Brosina H Consumer	United States	Los Angeles	California	90032	West	TEC-PH-10	Technology	Phones	Mitel 532i	907.152	6	0.2	90.7152
9	CA-2011-1	9/6/2011	14-06-2011	Standard	BH-11710	Brosina H Consumer	United States	Los Angeles	California	90032	West	OFF-BI-10	Office Supplies	Binders	DXL Angle	18.504	3	0.2	5.7825
10	CA-2011-1	9/6/2011	14-06-2011	Standard	BH-11710	Brosina H Consumer	United States	Los Angeles	California	90032	West	OFF-AP-1C	Office Supplies	Appliance	Belkin F5C	114.9	5	0	34.47
11	CA-2011-1	9/6/2011	14-06-2011	Standard	BH-11710	Brosina H Consumer	United States	Los Angeles	California	90032	West	FUR-TA-10	Furniture	Tables	Chromcra	1706.18	9	0.2	85.3092
12	CA-2011-1	9/6/2011	14-06-2011	Standard	BH-11710	Brosina H Consumer	United States	Los Angeles	California	90032	West	TEC-PH-10	Technology	Phones	Konftel 25	911.424	4	0.2	68.3568
13	CA-2014-1	16-04-2014	21-04-2014	Standard	AA-10480	Andrew Al Consumer	United States	Concord	North Carolina	28027	South	OFF-PA-1C	Office Supplies	Paper	Xerox 196	15.552	3	0.2	5.4432
14	CA-2013-1	6/12/2013	11/12/2013	Standard	IM-15070	Irene Mad Consumer	United States	Seattle	Washington	98103	West	OFF-BI-10	Office Supplies	Binders	Fellowes F	407.976	3	0.2	132.592
15	US-2012-1	122-11-2012	26-11-2012	Standard	HP-14815	Harold Pa Home Office	United States	Fort Worth	Texas	76106	Central	OFF-AP-1C	Office Supplies	Appliance	Holmes Re	68.81	5	0.8	-123.858
16	US-2012-1	122-11-2012	26-11-2012	Standard	HP-14815	Harold Pa Home Office	United States	Fort Worth	Texas	76106	Central	OFF-BI-10	Office Supplies	Binders	Storex Dur	2.544	3	0.8	-3.816
17	CA-2011-1	11/11/2011	18-11-2011	Standard	PK-19075	Pete Kriz Consumer	United States	Madison	Wisconsin	53711	Central	OFF-ST-10	Office Supplies	Storage	Stur-D-Sto	665.88	6	0	13.3176
18	CA-2011-1	113-05-2011	15-05-2011	Second Class	AG-10270	Alejandro Consumer	United States	West Jord	Utah	84084	West	OFF-ST-10	Office Supplies	Storage	Fellowes S	55.5	2	0	9.99
19	CA-2011-1	127-08-2011	1/9/2011	Second Class	ZD-21925	Zuschuss Consumer	United States	San Francisco	California	94109	West	OFF-AR-1C	Office Supplies	Art	Newell 34	8.56	2	0	2.4824
20	CA-2011-1	127-08-2011	1/9/2011	Second Class	ZD-21925	Zuschuss Consumer	United States	San Francisco	California	94109	West	TEC-PH-10	Technology	Phones	Cisco SPA	213.48	3	0.2	16.011
21	CA-2011-1	127-08-2011	1/9/2011	Second Class	ZD-21925	Zuschuss Consumer	United States	San Francisco	California	94109	West	OFF-BI-10	Office Supplies	Binders	Wilson Jo	22.72	4	0.2	7.384
22	CA-2013-1	10/12/2013	14-12-2013	Standard	KB-16585	Ken Black Corporate	United States	Fremont	Nebraska	68025	Central	OFF-AR-1C	Office Supplies	Art	Newell 31	19.46	7	0	5.0596
23	CA-2013-1	10/12/2013	14-12-2013	Standard	KB-16585	Ken Black Corporate	United States	Fremont	Nebraska	68025	Central	OFF-AP-1C	Office Supplies	Appliance	Acco Six-C	60.34	7	0	15.6884
24	US-2014-1	17-07-2014	19-07-2014	Second Class	SF-20065	Sandra Flt Consumer	United States	Philadelphia	Pennsylvania	19140	East	FUR-CH-1C	Furniture	Chairs	Global De	71.372	2	0.3	-1.0196

### Data Cleaning and Data Transformation:

1. Date Time formatting in dd/mm/yyyy format
2. Extract Order Year Month and Year
3. Calculate Delivery duration days
4. Calculate Total revenue after discount
5. Total Order
6. Total Quantity
7. Total Revenue
8. Average Sales
9. Average Duration

### KPIs:

1. Total Sales: \$10M
2. Total Quantity: 38k
3. Profit Margin: 0.03
4. Average Sales: \$990
5. Average Delivery Status: 3.96 days
6. Total Orders: 10k

### Visualization:

1. Total and Average Sales by Month and Year
2. Total and Average Sales by City and States
3. Total and Average Sales by Segment
4. Total and Average Sales by category and sub-category

## 5. Total and Average Sales by Product

### Slicers:

- Year
- Month
- Region
- Product
- Ship Mode

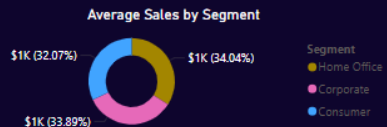
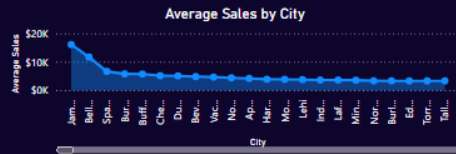
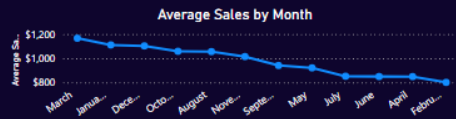
### Dashboard:



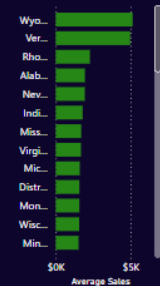


## SUPERSTORE SALES INSIGHTS

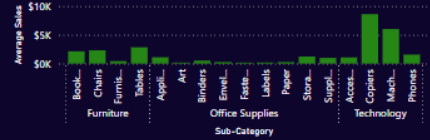
Total Revenue **\$10M** Total Quantity **38K** Profit Margin **0.03** Average Sales **\$990** Avg Delivery... **3.96** Total Order **10K**



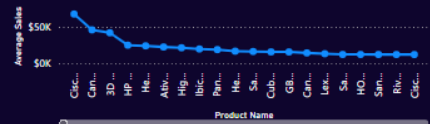
### Average Sales by State



### Average Sales by Category and Sub-Category



### Average Sales by Product Name



### KPIS

#### Measures

Average Sales

#### Year, Month

All

#### Ship Mode

All

#### Product Na...

All

#### Region

All