

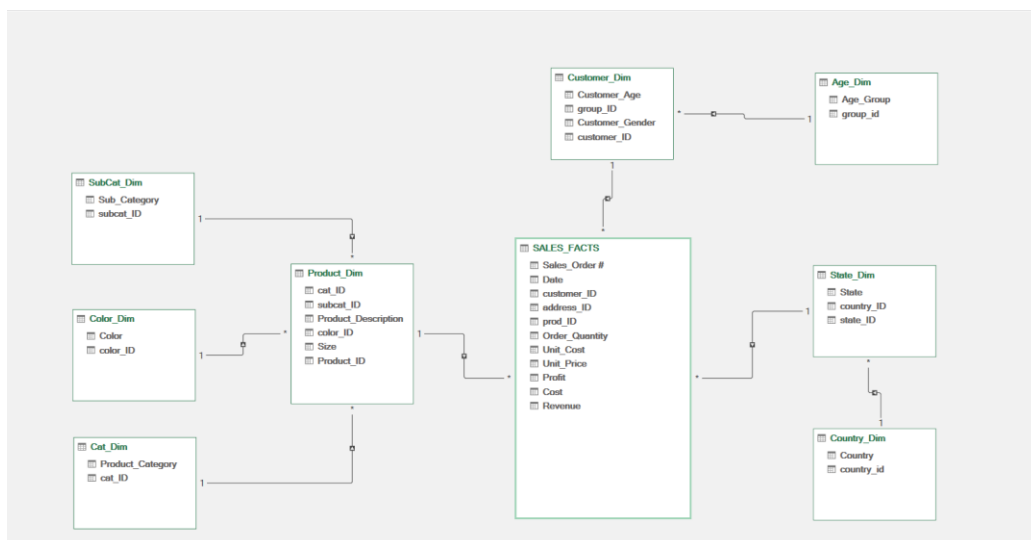
Manansala, Joshua N.

BSCS-C301

Midterm Lab Task # 1: Data Preparation and Cleaning

This practice task focuses on transforming a messy, unorganized dataset into a clean and structured format ready for professional analysis. It involves fixing text inconsistencies—such as removing extra spaces and correcting capitalization—while separating combined data fields like "Customer" and "State" into their own columns. The final goal is to normalize the data table.

Sales_Order #	Date	customer_ID	address_ID	prod_ID	Order_Quantity	Unit_Cost	Unit_Price	Profit	Cost	Revenue
000261695	01/12/2021	c-1031	s-115	p-1001	4	\$1,252.00	\$2,295.00	\$4,172.00	\$5,008.00	\$9,180.00
000261696	01/12/2021	c-1039	s-114	p-1001	1	\$1,266.00	\$2,320.00	\$1,054.00	\$1,266.00	\$2,320.00
000261697	02/12/2021	c-1027	s-115	p-1003	2	\$420.00	\$769.00	\$698.00	\$840.00	\$1,538.00
000261698	02/12/2021	c-1015	s-101	p-1003	1	\$420.00	\$769.00	\$349.00	\$420.00	\$769.00
000261699	03/12/2021	c-1027	s-115	p-1001	2	\$1,252.00	\$2,295.00	\$2,086.00	\$2,504.00	\$4,590.00
000261700	03/12/2021	c-1005	s-114	p-1001	1	\$1,252.00	\$2,295.00	\$1,043.00	\$1,252.00	\$2,295.00
000261701	03/12/2021	c-1027	s-116	p-1001	1	\$1,252.00	\$2,295.00	\$1,043.00	\$1,252.00	\$2,295.00
000261702	04/12/2021	c-1015	s-101	p-1003	4	\$420.00	\$769.00	\$1,396.00	\$1,680.00	\$3,076.00
000261703	05/12/2021	c-1031	s-115	p-1001	4	\$1,252.00	\$2,295.00	\$4,172.00	\$5,008.00	\$9,180.00
000261704	05/12/2021	c-1037	s-111	p-1001	4	\$1,252.00	\$2,295.00	\$4,172.00	\$5,008.00	\$9,180.00
000261705	05/12/2021	c-1022	s-102	p-1001	1	\$1,266.00	\$2,320.00	\$1,054.00	\$1,266.00	\$2,320.00
000261706	05/12/2021	c-1027	s-115	p-1001	1	\$1,252.00	\$2,295.00	\$1,043.00	\$1,252.00	\$2,295.00
000261707	06/12/2021	c-1004	s-114	p-1003	3	\$420.00	\$769.00	\$1,047.00	\$1,260.00	\$2,307.00
000261708	06/12/2021	c-1009	s-105	p-1001	1	\$1,252.00	\$2,295.00	\$1,043.00	\$1,252.00	\$2,295.00
000261709	06/12/2021	c-1024	s-101	p-1001	1	\$1,252.00	\$2,295.00	\$1,043.00	\$1,252.00	\$2,295.00
000261710	06/12/2021	c-1041	s-114	p-1001	1	\$1,266.00	\$2,320.00	\$1,054.00	\$1,266.00	\$2,320.00
000261711	07/12/2021	c-1013	s-115	p-1003	4	\$420.00	\$769.00	\$1,396.00	\$1,680.00	\$3,076.00
000261712	07/12/2021	c-1029	s-115	p-1001	2	\$1,266.00	\$2,320.00	\$2,108.00	\$2,532.00	\$4,640.00
000261713	08/12/2021	c-1003	s-101	p-1009	4	\$308.00	\$565.00	\$1,028.00	\$1,232.00	\$2,260.00
000261714	08/12/2021	c-1013	s-105	p-1001	4	\$1,266.00	\$2,320.00	\$4,216.00	\$5,064.00	\$9,280.00
000261715	08/12/2021	c-1031	s-117	p-1001	2	\$1,252.00	\$2,295.00	\$2,086.00	\$2,504.00	\$4,590.00
000261716	08/12/2021	c-1022	s-115	p-1009	4	\$295.00	\$540.00	\$245.00	\$1,180.00	\$2,160.00
000261717	09/12/2021	c-1019	s-103	p-1011	2	\$1,898.00	\$3,375.00	\$2,954.00	\$3,796.00	\$6,750.00
000261718	09/12/2021	c-1035	s-112	p-1001	1	\$1,266.00	\$2,320.00	\$1,054.00	\$1,266.00	\$2,320.00
000261719	10/12/2021	c-1020	s-115	p-1001	2	\$1,252.00	\$2,295.00	\$2,086.00	\$2,504.00	\$4,590.00
000261720	10/12/2021	c-1033	s-101	p-1001	2	\$1,252.00	\$2,295.00	\$2,086.00	\$2,504.00	\$4,590.00
000261721	10/12/2021	c-1007	s-114	p-1001	1	\$1,252.00	\$2,295.00	\$1,043.00	\$1,252.00	\$2,295.00
000261722	10/12/2021	c-1020	s-115	p-1009	1	\$295.00	\$540.00	\$245.00	\$295.00	\$540.00
000261723	10/12/2021	c-1020	s-116	p-1011	1	\$1,912.00	\$3,400.00	\$1,488.00	\$1,912.00	\$3,400.00
000261724	10/12/2021	c-1029	s-101	p-1001	1	\$1,252.00	\$2,295.00	\$1,043.00	\$1,252.00	\$2,295.00
000261725	11/12/2021	c-1005	s-106	p-1001	3	\$1,252.00	\$2,295.00	\$3,129.00	\$3,756.00	\$6,885.00
000261726	11/12/2021	c-1035	s-101	p-1003	2	\$420.00	\$769.00	\$698.00	\$840.00	\$1,538.00
000261727	11/12/2021	c-1008	s-105	p-1001	1	\$1,266.00	\$2,320.00	\$1,043.00	\$1,266.00	\$2,320.00



	A	B	C	D
1	Customer_Age	group_ID	Customer_Gender	customer_ID
2	17	g-01	Male	c-1001
3	18	g-01	Male	c-1002
4	19	g-01	Female	c-1003
5	23	g-01	Male	c-1004
6	24	g-01	Female	c-1005
7	25	g-02	Male	c-1006
8	26	g-02	Male	c-1007
9	26	g-02	Female	c-1008
10	27	g-02	Male	c-1009
11	27	g-02	Female	c-1010
12	29	g-02	Female	c-1011
13	29	g-02	Male	c-1012
14	30	g-02	Male	c-1013
15	30	g-02	Female	c-1014
16	31	g-02	Female	c-1015
17	31	g-02	Male	c-1016
18	32	g-02	Female	c-1017
19	32	g-02	Male	c-1018
20	33	g-02	Female	c-1019
21	34	g-02	Female	c-1020
22	34	g-02	Male	c-1021
23	35	g-03	Female	c-1022
24	35	g-03	Male	c-1023
25	36	g-01	Male	c-1024
26	36	g-03	Female	c-1025
27	36	g-03	Male	c-1026
28	37	g-03	Male	c-1027
29	37	g-03	Female	c-1028
30	38	g-03	Male	c-1029
31	38	g-03	Female	c-1030
32	39	g-03	Female	c-1031
33	39	g-03	Male	c-1032
34	40	g-03	Male	c-1033

Navigation: < > BikeSales_FACTS Customer Age_group Product Prod_categ

	A	B
1	Age_Group	group_id
2	Youth (<25)	g-01
3	Young Adults (25-34)	g-02
4	Adults (35-64)	g-03
5		

	A	B
1	Product_Category	cat_ID
2	Bikes	cat-01
3		

	A	B
1	Sub_Category	subcat_ID
2	Mountain Bikes	sc-01
3		

	A	B
1	Color	color_ID
2	Black	col-101
3	Silver	col-102

	A	B	C	D	E	F
1	cat_ID	subcat_ID	Product_Description	color_ID	Size	Product_ID
2	cat-01	sc-01	Mountain-200	col-101		46 p-1001
3	cat-01	sc-01	Mountain-200	col-102		42 p-1002
4	cat-01	sc-01	Mountain-400-W	col-102		46 p-1003
5	cat-01	sc-01	Mountain-400-W	col-102		42 p-1004
6	cat-01	sc-01	Mountain-200	col-101		38 p-1005
7	cat-01	sc-01	Mountain-200	col-102		38 p-1006
8	cat-01	sc-01	Mountain-200	col-101		42 p-1007
9	cat-01	sc-01	Mountain-400-W	col-102		38 p-1008
10	cat-01	sc-01	Mountain-500	col-102		42 p-1009
11	cat-01	sc-01	Mountain-500	col-101		42 p-1010
12	cat-01	sc-01	Mountain-100	col-101		38 p-1011
13	cat-01	sc-01	Mountain-500	col-101		40 p-1012
14	cat-01	sc-01	Mountain-100	col-102		44 p-1013
15	cat-01	sc-01	Mountain-500	col-102		40 p-1014
16	cat-01	sc-01	Mountain-200	col-102		46 p-1015
17	cat-01	sc-01	Mountain-500	col-101		44 p-1016
18	cat-01	sc-01	Mountain-100	col-101		48 p-1017
19	cat-01	sc-01	Mountain-500	col-101		52 p-1018

	A	B	C
1	State	country_ID	state_ID
2	New South Wales	cn-101	s-101
3	Queensland	cn-101	s-102
4	Victoria	cn-101	s-103
5	South Australia	cn-101	s-104
6	British Columbia	cn-102	s-105
7	Seine (Paris)	cn-103	s-106
8	Seine et Marne	cn-103	s-107
9	Seine Saint Denis	cn-103	s-108
10	Nord	cn-103	s-109
11	Somme	cn-103	s-110
12	Nordrhein-Westfalen	cn-104	s-111
13	Hamburg	cn-104	s-112
14	Hessen	cn-104	s-113
15	England	cn-105	s-114
16	California	cn-106	s-115
17	Washington	cn-106	s-116
18	Oregon	cn-106	s-117

	A	B
1	Country	country_id
2	australia	cn-101
3	canada	cn-102
4	france	cn-103
5	germany	cn-104
6	united kingdom	cn-105
7	united states	cn-106