

GROUP 1 MVP TRAINING MANUAL

Step by Step guide to get you ready for using the MVP.

Cumulated Efforts
by:
Rick Kalouche
Zoe Fong
Michelle Wang
Christopher Rigor
Thomas Yap

Group 1 Training Manual

Contents

<i>Data Cleansing</i>	2
<i>Process of cleansing and preparing data</i>	2
<i>Cleansing and preparing a data set</i>	2
<i>Integrating separate data sets</i>	5
<i>Windows Installation Guide for Anaconda</i>	12
<i>Jupyter Access Instructions</i>	19
<i>Packages</i>	22
<i>Dataset and Dataframe Manipulation</i>	22
<i>Data Cleaning and Manipulation</i>	22
<i>Aggregating Data</i>	23
<i>Documentation</i>	24
<i>Legend</i>	24
<i>Packages</i>	24
<i>Dataset and Dataframe Manipulation</i>	25
<i>Data Cleaning and Manipulation</i>	25
<i>Aggregating Data</i>	26
<i>Data Visualisation / Exploration</i>	26
<i>Modelling</i>	28
<i>Creating Model Dataframe</i>	28
<i>KNN</i>	30
<i>ARIMA</i>	30
<i>DTR</i>	32
<i>Logistic Regression</i>	33
<i>Prediction</i>	34
<i>Choosing Parameters</i>	34

Group 1 Training Manual

Data Cleansing

The cleansing of the data sets before they are imported into Jupyter is vital to ensure that the program runs smoothly free of errors in order to give accurate results. This is especially important when dealing with large amounts of data to be able to take advantage of running effective analysis. This process will involve identifying inconsistencies, mistakes and errors in the data and fixing these issues, as well as reformatting and cutting. The only tool required for data cleaning is Microsoft Excel.

The main outcome of this process is to create high quality data. Meeting the goal of quality data ensures that the data being used is thoroughly cleansed. Quality data is defined by five characteristics which are:

- Accuracy - The information that the data contains is correct.
- Reliability - The data doesn't contradict other pieces of data in other data sets.
- Completeness - All elements of the data are readily available.
- Relevance - The data collected is suitable and appropriate for its intended purpose.
- Timeliness - The data is up to date and accurate to a reasonable time period.

Process of cleansing and preparing data

This section aims to explain how to cleanse, prepare and integrate raw data sets using Excel to ensure it's ready for analysis through Jupyter. Additional instructions will be provided for users who want to combine and integrate separate data sets from steps X to Y.

Cleansing and preparing a data set

This section shows how to cleanse and prepare a data set for analysis step by step.

Step 1: Data inspection

Go through the data set and remove duplicated or irrelevant columns. Be aware of the primary keys, they are the unique identifier/value in the data set. Do not delete the primary key.

In the example below, the orange boxes represent the column that needs to be removed.

- “SHIPYEAR” and “SHIPEMONTH” are duplicated and getting deleted as the year and month are already mentioned in “SHIPDATE”.
- “C_STATE” is irrelevant and getting removed as “C_ZIP” is provided.
- “DESCR”, which is the description of the item is irrelevant and getting removed.

Group 1 Training Manual

Screenshot of Microsoft Excel showing a table titled "Petwise - PW_ConsolidatedSales_October_2020". The table contains data from columns A to N, with rows 1 through 38 visible. Column H is highlighted with a yellow border.

VENDOR	SHIPYEAR	SHIPMONTH	SHIPTDATE	ORDERKEY	CONSIGNEEKEY	C_COMPANY	C_STATE	C_ZIP	EXTERNOERKEY	SKU	DESCR	ORIGINALQTY	SHIPPEDQTY
1 BAYER	2020	October	2/0/20	277251	500418	PETBARN P/L TOOOWOMBA	QLD	4350	POR1371478	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1
2 BAYER	2020	October	2/0/20	277305	503669	PETBARN P/L MITCHELTON	QLD	4053	POR1371516	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1
3 BAYER	2020	October	2/0/20	277313	502714	PETBARN P/L MAROCHYDORIE NEW	QLD	4558	POR1371523	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1
4 BAYER	2020	October	2/0/20	278572	504531	Good Price Pharmacy Warehouse Cannon Hill	QLD	4170	145157	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1
5 BAYER	2020	October	2/0/20	278663	500347	PETBARN P/L BUNDALL	QLD	4217	POR1372599	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	2	2
6 BAYER	2020	October	6/10/20	277933	501392	CITY FARMERS ROCKINGHAM	WA	6168	POR1371934	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1
7 BAYER	2020	October	7/10/20	278666	500355	PETBARN P/L CHARMHAVEN	NSW	2263	POR1372617	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1
8 BAYER	2020	October	7/10/20	281190	503722	PETBARN P/L BOX HILL	VIC	3128	POR1374070	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1
9 BAYER	2020	October	8/10/20	281376	500391	PETBARN P/L MILE END	SA	5031	POR1373373	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1
10 BAYER	2020	October	9/10/20	281156	500354	PETBARN P/L CHADSTONE	VIC	3148	POR1373961	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	2	2
11 BAYER	2020	October	9/10/20	281184	501387	PETBARN P/L NARRE WARREN	VIC	3802	POR1374028	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1
12 BAYER	2020	October	9/10/20	281826	503659	PETBARN P/L FRANKSTON	WA	6155	POR1374462	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1
13 BAYER	2020	October	9/10/20	282667	503613	PETBARN P/L WENDOURREE	VIC	3355	POR1371973	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1
14 BAYER	2020	October	10/10/20	281847	500355	PETBARN P/L KOCOMA	QLD	4053	POR1374457	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1
15 BAYER	2020	October	12/0/20	281807	501392	CITY FARMERS DOTHFIELD	WA	6058	POR1374064	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1
16 BAYER	2020	October	12/0/20	282698	500415	PETBARN P/L SOUTH STRATHFIELD	NSW	2135	POR1374004	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1
17 BAYER	2020	October	12/0/20	282843	503611	PETBARN P/L ST KILDA	VIC	3182	POR1371527	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	2	2
18 BAYER	2020	October	13/0/20	282493	504363	PETBARN P/L TOOMBUL	QLD	4012	POR1375045	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1
19 BAYER	2020	October	13/0/20	282701	500341	PETBARN P/L BAYSWATER	VIC	3153	POR1374013	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1
20 BAYER	2020	October	13/0/20	282842	503659	PETBARN P/L FRANKSTON	VIC	3199	POR1371518	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1
21 BAYER	2020	October	13/0/20	284037	500383	PETBARN P/L LAWNTON	QLD	4501	POR1375448	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1
22 BAYER	2020	October	13/0/20	284356	500359	PETBARN P/L CLAYTON	VIC	3168	POR1376025	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	2	2
23 BAYER	2020	October	13/10/20	284743	503927	PETBARN P/L MT GAMBIER	SA	5259	POR1376090	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1
24 BAYER	2020	October	14/10/20	283998	500431	PETBARN P/L WOLLONGONG	NSW	2506	POR1375328	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1
25 BAYER	2020	October	14/10/20	284022	503612	PETBARN P/L TWEED HEAD	QLD	4211	POR1375406	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1
26 BAYER	2020	October	14/10/20	284023	503609	PETBARN P/L NERANG	NSW	2289	POR1375464	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1
27 BAYER	2020	October	14/10/20	284042	500424	PETBARN P/L WATERGARDENS	VIC	3038	POR1375176	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1
28 BAYER	2020	October	14/10/20	285319	500354	PETBARN P/L CHADSTONE	VIC	3148	POR1376742	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	2	2
29 BAYER	2020	October	14/10/20	285349	501387	PETBARN P/L NARRE WARREN	VIC	3805	POR1376829	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	2	2
30 BAYER	2020	October	14/10/20	285358	504004	PETBARN P/L INDOOROOPIPOLY	QLD	4068	POR1376885	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1
31 BAYER	2020	October	14/10/20	285365	504050	PETBARN P/L WOOLLOONGABBA	QLD	4102	POR1376900	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1
32 BAYER	2020	October	14/10/20	285366	504485	PETBARN P/L PALMERSTON	NT	830	POR1376905	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1
33 BAYER	2020	October	14/10/20	285368	503688	PETBARN P/L SOUTH YARRA	NT	810	POR1376911	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1
34 BAYER	2020	October	15/10/20	284726	503688	PETBARN P/L SOUTH YARRA	VIC	3141	POR1376078	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1
35 BAYER	2020	October	15/10/20	285311	500356	PETBARN P/L CHATSWOOD	NSW	2067	POR1376726	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1
36 BAYER	2020	October	15/10/20	285325	500422	PETBARN P/L VIRGINIA	QLD	4014	POR1376756	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1

In order to delete, users are to select the column using the cursor, right click, and then select “delete”, as shown on the screenshot below.

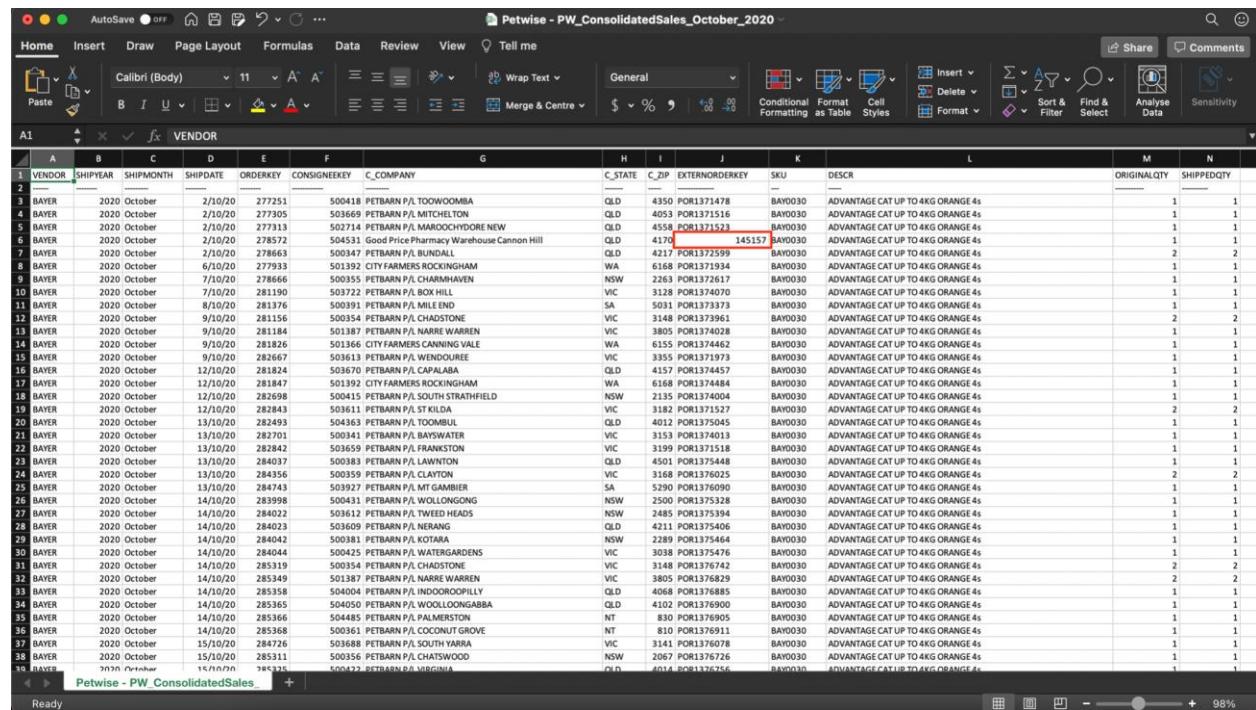
Screenshot of Microsoft Excel showing the same table as above, but with the "Delete" option highlighted in green in the context menu for column E. The "Delete" option is the second item in the list.

VENDOR	SHIPYEAR	SHIPMONTH	SHIPTDATE	DERKEY	CONSIGNEEKEY	C_COMPANY	C_STATE	C_ZIP	EXTERNOERKEY	SKU	DESCR	ORIGINALQTY	SHIPPEDQTY
1 BAYER	2020	October	2/0/20	277251	500418	PETBARN P/L TOOOWOMBA	QLD	4350	POR1371478	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1
2 BAYER	2020	October	2/0/20	277305	503669	PETBARN P/L MITCHELTON	QLD	4053	POR1371516	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1
3 BAYER	2020	October	2/0/20	277313	502714	PETBARN P/L MAROCHYDORIE NEW	QLD	4558	POR1371523	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1
4 BAYER	2020	October	2/0/20	278572	504531	Good Price Pharmacy Warehouse Cannon Hill	QLD	4170	145157	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1
5 BAYER	2020	October	2/0/20	278663	500347	PETBARN P/L BUNDALL	QLD	4217	POR1372599	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	2	2
6 BAYER	2020	October	2/0/20	281190	503722	PETBARN P/L BOX HILL	VIC	3128	POR1374070	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1
7 BAYER	2020	October	8/10/20	281376	500391	PETBARN P/L MILE END	SA	5031	POR1373373	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	2	2
8 BAYER	2020	October	9/10/20	281156	502714	PETBARN P/L CHADSTONE	VIC	3148	POR1373961	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1
9 BAYER	2020	October	9/10/20	281184	501387	PETBARN P/L NARRE WARREN	VIC	3802	POR1374028	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1
10 BAYER	2020	October	9/10/20	281826	503613	CITY FARMERS CANNING VALE	WA	6155	POR1374462	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1
11 BAYER	2020	October	10/10/20	281847	503927	PETBARN P/L MT GAMBIER	SA	5259	POR1376090	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1
12 BAYER	2020	October	14/10/20	283998	500431	PETBARN P/L WOLLONGONG	NSW	2506	POR1375328	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1
13 BAYER	2020	October	14/10/20	284022	503612	PETBARN P/L TWEED HEAD	NSW	2451	POR1375394	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1
14 BAYER	2020	October	14/10/20	284023	503609	PETBARN P/L NERANG	QLD	4211	POR1375406	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1
15 BAYER	2020	October	14/10/20	284044	500425	PETBARN P/L WATERGARDENS	VIC	3038	POR1375476	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1
16 BAYER	2020	October	14/10/20	285319	500354	PETBARN P/L CHADSTONE	VIC	3148	POR1376742	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	2	2
17 BAYER	2020	October	14/10/20	285349	501387	PETBARN P/L NARRE WARREN	VIC	3805	POR1376829	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	2	2
18 BAYER	2020	October	14/10/20	285358	504004	PETBARN P/L INDOOROOPIPOLY	QLD	4068	POR1376885	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1
19 BAYER	2020	October	14/10/20	285365	504050	PETBARN P/L WOOLLOONGABBA	QLD	4102	POR1376900	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1
20 BAYER	2020	October	14/10/20	285366	504485	PETBARN P/L PALMERSTON	NT	830	POR1376905	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1
21 BAYER	2020	October	14/10/20	285368	500361	PETBARN P/L COCONUT GROVE	NT	810	POR1376911	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1
22 BAYER	2020	October	15/10/20	284726	503688	PETBARN P/L SOUTH YARRA	VIC	3141	POR1376078	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1
23 BAYER	2020	October	15/10/20	285311	500356	PETBARN P/L CHATSWOOD	NSW	2067	POR1376726	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1
24 BAYER	2020	October	15/10/20	285325	500422	PETBARN P/L VIRGINIA	QLD	4014	POR1376756	BAY030	ADVANTAGE CAT UP TO 4KG ORANGE 4s	1	1

Group 1 Training Manual

Step 2: Structural errors correction

After step 1, go through the data set again to find any typos or syntax error. They could be mis-spelling words, capitalisation, etc. The red box in the following example represents the errors. Fix the error if they are minor. If they cannot be fixed, one can move to the next step.



The screenshot shows a Microsoft Excel spreadsheet titled "Petwise - PW_ConsolidatedSales_October_2020". The data is organized into several columns: VENDOR, SHIPYEAR, SHIPMONTH, SHIPDATE, ORDERKEY, CONSIGNEEKEY, C_COMPANY, C_STATE, C_ZIP, EXTERNOERDERKEY, SKU, DESCR, ORIGINALQTY, and SHIPPEDQTY. A red box highlights the value "145157" in the EXTERNOERDERKEY column for row 7, which corresponds to a row where the C_COMPANY value is partially cut off ("CITY FARMERS ROCKINGHAM"). Another red box highlights the value "145157" in the EXTERNOERDERKEY column for row 14, which corresponds to a row where the C_COMPANY value is partially cut off ("CITY FARMERS CANNING VALE"). The table has 39 rows, with the last row being a summary or header row.

Step 3: Take notes of the missing data

After step 2, if there is missing data, the whole column should be dropped or import data based on observations on the rest of the data. This may involve making up the missing data.

In the following example, two data sets are shown, the one on the left is data from October, the right is from November. Two data sets are meant to be in the same format, however, a column is missing in the October one. Users will have to decide whether they want to drop the corresponding column in other data sets or make the data up according to the data. In this case, it is better to drop the column in the November set.

Group 1 Training Manual

The screenshot shows two Microsoft Excel windows side-by-side. Both windows have the title bar 'Western Pet - BI Sales Report Octo...' and 'Western - BI Sales Report Novem...'. The left window (October) has a column header 'Quantity' with a red box around it and an orange arrow pointing to the right window (November). The right window (November) has a column header 'STAFF PURCHASES' with a red box around it. Both windows show a grid of data with columns for Product ID, Description, Quantity, Price, and Net Total.

Integrating separate data sets

The following steps are for users who want to combine separate data sets into one. For example, three different data sets containing the sales of a product are separated into January, February and March sales data sets. The user may want to combine these data sets for analysis. The following steps will explain how to integrate these months into one data set.

Step 4: Combining data sets

First, ensure the data sets have the same columns and data types before combining. In the screenshots below, Western sales reports for October and November have the same exact columns and data types. Therefore they are ready for integration.

Group 1 Training Manual

Western - BI Sales Report November 2020

Sheet1

A1	Stockcode	Suppliercode	Barcode	Description	Quantity	F	Price	Acc. Name	H	I	J	K
									Postcode			
2	CHA0010	84266825	9310160821720	ADVANTAGE SMALL CAT 0.4KG 1PK	1	11.97	tasmanian animal supplies		7325			
3	CHA0010	84266825	9310160821720	ADVANTAGE SMALL CAT 0.4KG 1PK	3	11.97	kaba deals "use account 4088"		3216			
4	CHA0010	84266825	9310160821720	ADVANTAGE SMALL CAT 0.4KG 1PK	3	11.97	taree pet barn		2430			
5	CHA0010	84266825	9310160821720	ADVANTAGE SMALL CAT 0.4KG 1PK	1	11.97	kellyville pets		2155			
6	CHA0010	84266825	9310160821720	ADVANTAGE SMALL CAT 0.4KG 1PK	1	11.97	woonona petfood & produce pty ltd		2518			
7	CHA0010	84266825	9310160821720	ADVANTAGE SMALL CAT 0.4KG 1PK	1	11.97	petstock traralgon		3844			
8	CHA0010	84266825	9310160821720	ADVANTAGE SMALL CAT 0.4KG 1PK	2	11.97	petstock ormond		3204			
9	CHA0010	84266825	9310160821720	ADVANTAGE SMALL CAT 0.4KG 1PK	1	11.97	petstock granbourne		3977			
10	CHA0010	84266825	9310160821720	ADVANTAGE SMALL CAT 0.4KG 1PK	1	11.97	petstock traralgon		3844			
11	CHA0010	84266825	9310160821720	ADVANTAGE SMALL CAT 0.4KG 1PK	2	11.97	taree pet barn		2430			
12	CHA0010	84266825	9310160821720	ADVANTAGE SMALL CAT 0.4KG 1PK	1	11.97	woonona petfood & produce pty ltd		2518			
13	CHA0010	84266825	9310160821720	ADVANTAGE SMALL CAT 0.4KG 1PK	2	11.97	petstock mildura		3500			
14	CHA0010	84266825	9310160821720	ADVANTAGE SMALL CAT 0.4KG 1PK	2	11.97	petstock rosebud		3940			
15	CHA0010	84266825	9310160821720	ADVANTAGE SMALL CAT 0.4KG 1PK	1	11.97	woonona petfood & produce pty ltd		2518			
16	CHA0010	84266825	9310160821720	ADVANTAGE SMALL CAT 0.4KG 1PK	2	11.97	petstock port lincoln		5606			
17	CHA0010	84266825	9310160821720	ADVANTAGE SMALL CAT 0.4KG 1PK	4	11.97	taree pet barn		2430			
18	CHA0010	84266825	9310160821720	ADVANTAGE SMALL CAT 0.4KG 1PK	5	11.97	petstock traralgon		3844			
19	CHA0010	84266825	9310160821720	ADVANTAGE SMALL CAT 0.4KG 1PK	3	11.97	petstock granbourne		3977			
20	CHA0010	84266825	9310160821720	ADVANTAGE SMALL CAT 0.4KG 1PK	2	11.97	petstock ormond		3204			
21	CHA0010	84266825	9310160821720	ADVANTAGE SMALL CAT 0.4KG 1PK	5	11.97	petstock melton		3337			
22	CHA0010	84266825	9310160821720	ADVANTAGE SMALL CAT 0.4KG 1PK	1	11.97	petstock port lincoln		6606			
23	CHA0200	4506646	9310160857787	ADVANTAGE SMALL CAT 0.4KG 4PACK	3	38.44	poplar petfood & produce		2526			
24	CHA0200	4506646	9310160857787	ADVANTAGE SMALL CAT 0.4KG 4PACK	1	38.44	pet depot		2040			
25	CHA0200	4506646	9310160857787	ADVANTAGE SMALL CAT 0.4KG 4PACK	1	38.44	pet o ashfield		2131			
26	CHA0200	4506646	9310160857787	ADVANTAGE SMALL CAT 0.4KG 4PACK	2	38.44	pet o northmead		2152			
27	CHA0200	4506646	9310160857787	ADVANTAGE SMALL CAT 0.4KG 4PACK	2	38.44	pet o annandale		2038			
28	CHA0200	4506646	9310160857787	ADVANTAGE SMALL CAT 0.4KG 4PACK	1	38.44	taree pet barn		2430			
29	CHA0200	4506646	9310160857787	ADVANTAGE SMALL CAT 0.4KG 4PACK	2	38.44	aldinga one stop pet shop & aquaria		5173			
30	CHA0200	4506646	9310160857787	ADVANTAGE SMALL CAT 0.4KG 4PACK	3	38.44	petshop discount drugstore		2229			
31	CHA0200	4506646	9310160857787	ADVANTAGE SMALL CAT 0.4KG 4PACK	1	38.44	petstock traralgon		3844			
32	CHA0200	4506646	9310160857787	ADVANTAGE SMALL CAT 0.4KG 4PACK	1	38.44	pittwater pet		2102			
33	CHA0200	4506646	9310160857787	ADVANTAGE SMALL CAT 0.4KG 4PACK	2	38.44	poplar petfood & produce		2526			
34	CHA0200	4506646	9310160857787	ADVANTAGE SMALL CAT 0.4KG 4PACK	1	38.44	pet o beverly hills		2209			
35	CHA0200	4506646	9310160857787	ADVANTAGE SMALL CAT 0.4KG 4PACK	1	38.44	pet o alexandria		2015			
36	CHA0200	4506646	9310160857787	ADVANTAGE SMALL CAT 0.4KG 4PACK	1	38.44	pet o mona vale		2103			
37	CHA0200	4506646	9310160857787	ADVANTAGE SMALL CAT 0.4KG 4PACK	1	38.44	pet o annandale		2038			
38	CHA0200	4506646	9310160857787	ADVANTAGE SMALL CAT 0.4KG 4PACK	1	38.44	kellyville pets		2155			
39	CHA0200	4506646	9310160857787	ADVANTAGE SMALL CAT 0.4KG 4PACK	1	38.44	taree pet barn		2430			
40	CHA0200	4506646	9310160857787	ADVANTAGE SMALL CAT 0.4KG 4PACK	1	38.44	petstock traralgon		3844			
41	CHA0200	4506646	9310160857787	ADVANTAGE SMALL CAT 0.4KG 4PACK	2	38.44	pet o ashfield		2131			
42	CHA0200	4506646	9310160857787	ADVANTAGE SMALL CAT 0.4KG 4PACK	4	38.44	aldinga one stop pet shop		2229			

Sheet1

H1	Stockcode	Suppliercode	Barcode	Description	Quantity	F	Price	Acc. Name	H	I	J	K
									Postcode			
2	CHA0010	84266825	9310160821720	ADVANTAGE SMALL CAT 0.4KG 1PK	1	11.97	TAREE PET BARN		2430			
3	CHA0010	84266825	9310160821720	ADVANTAGE SMALL CAT 0.4KG 1PK	8	11.97	PETSTOCK TRARALGON		3844			
4	CHA0010	84266825	9310160821720	ADVANTAGE SMALL CAT 0.4KG 1PK	1	11.97	DAKEY RURAL SUPPLIES		4401			
5	CHA0010	84266825	9310160821720	ADVANTAGE SMALL CAT 0.4KG 1PK	1	11.97	PETSTOCK ROSEBUD		3940			
6	CHA0010	84266825	9310160821720	ADVANTAGE SMALL CAT 0.4KG 1PK	2	11.97	PETSTOCK PORT LINCOLN		5606			
7	CHA0010	84266825	9310160821720	ADVANTAGE SMALL CAT 0.4KG 1PK	2	11.97	PETSTOCK CRANBOURNE		3977			
8	CHA0010	84266825	9310160821720	ADVANTAGE SMALL CAT 0.4KG 1PK	2	11.97	ALDINGA ONE STOP PET SHOP & AQUARIUMS		5173			
9	CHA0010	84266825	9310160821720	ADVANTAGE SMALL CAT 0.4KG 1PK	1	11.97	PETSTOCK ROSEBUD		3940			
10	CHA0010	84266825	9310160821720	ADVANTAGE SMALL CAT 0.4KG 1PK	1	11.97	PETSTOCK PORT LINCOLN		5205			
11	CHA0010	84266825	9310160821720	ADVANTAGE SMALL CAT 0.4KG 1PK	1	11.97	PETSTOCK ORMOND		3204			
12	CHA0010	84266825	9310160821720	ADVANTAGE SMALL CAT 0.4KG 1PK	1	11.97	PETSTOCK ORMOND		3204			
13	CHA0010	84266825	9310160821720	ADVANTAGE SMALL CAT 0.4KG 1PK	1	11.97	KELLYVILLE PETS		2155			
14	CHA0010	84266825	9310160821720	ADVANTAGE SMALL CAT 0.4KG 1PK	1	11.97	TAREE PET BARN		2430			
15	CHA0200	4506646	9310160857787	ADVANTAGE SMALL CAT 0.4KG 4PACK	2	38.44	PET O ANNADALE		2038			
16	CHA0200	4506646	9310160857787	ADVANTAGE SMALL CAT 0.4KG 4PACK	1	38.44	PETSTOCK TRARALGON		3844			
17	CHA0200	4506646	9310160857787	ADVANTAGE SMALL CAT 0.4KG 4PACK	1	38.44	PET O HORNSBY		2077			
18	CHA0200	4506646	9310160857787	ADVANTAGE SMALL CAT 0.4KG 4PACK	1	38.44	EDGECLIFF PETS (THE CATS WHISKERS T/A)		2027			
19	CHA0200	4506646	9310160857787	ADVANTAGE SMALL CAT 0.4KG 4PACK	1	38.44	BEACHMERE DISCOUNT DRUG STORE		4510			
20	CHA0200	4506646	9310160857787	ADVANTAGE SMALL CAT 0.4KG 4PACK	2	38.44	PET CARE 2000 CARINGBAH		2229			
21	CHA0200	4506646	9310160857787	ADVANTAGE SMALL CAT 0.4KG 4PACK	1	38.44	PET STORE		2232			
22	CHA0200	4506646	9310160857787	ADVANTAGE SMALL CAT 0.4KG 4PACK	1	38.44	PET O ASHFIELD		2131			
23	CHA0200	4506646	9310160857787	ADVANTAGE SMALL CAT 0.4KG 4PACK	1	38.44	PET O HORNSBY		2077			
24	CHA0200	4506646	9310160857787	ADVANTAGE SMALL CAT 0.4KG 4PACK	2	38.44	PETCARE 2000 YAGOONA		2199			
25	CHA0200	4506646	9310160857787	ADVANTAGE SMALL CAT 0.4KG 4PACK	1	38.44	CONDELL PARK PRODUCE		2200			
26	CHA0200	4506646	9310160857787	ADVANTAGE SMALL CAT 0.4KG 4PACK	1	38.44	PET O ASHFIELD		2131			
27	CHA0200	4506646	9310160857787	ADVANTAGE SMALL CAT 0.4KG 4PACK	2	38.44	IPSWICH PET & AQUARIUM		4305			
28	CHA0200	4506646	9310160857787	ADVANTAGE SMALL CAT 0.4KG 4PACK	1	38.44	PET O ALEXANDRIA		2015			
29	CHA0200	4506646	9310160857787	ADVANTAGE SMALL CAT 0.4KG 4PACK	1	38.44	PET O NORTHEMAD		2152			
30	CHA0200	4506646	9310160857787	ADVANTAGE SMALL CAT 0.4KG 4PACK	1	38.44	PET O BEVERLY HILLS		2038			
31	CHA0200	4506646	9310160857787	ADVANTAGE SMALL CAT 0.4KG 4PACK	2	38.44	ALDINGA ONE STOP PET SHOP & AQUARIUMS		5173			
32	CHA0200	4506646	9310160857787	ADVANTAGE SMALL CAT 0.4KG 4PACK	1	38.44	KELLYVILLE PETS		2155			
33	CHA0200	4506646	9310160857787	ADVANTAGE SMALL CAT 0.4KG 4PACK	1	38.44	PET O ANNADALE		2038			
34	CHA0200	4506646	9310160857787	ADVANTAGE SMALL CAT 0.4KG 4PACK	1	38.44	ANIMAL HOUSE BALLINA		2478			
35	CHA0200	4506646	9310160857787	ADVANTAGE SMALL CAT 0.4KG 4PACK	2	38.44	PETSTOCK TRARALGON		3844			
36	CHA0200	4506646	9310160857787	ADVANTAGE SMALL CAT 0.4KG 4PACK	1	38.44	PET O ALEXANDRIA		2015			
37	CHA0200	4506646	9310160857787	ADVANTAGE SMALL CAT 0.4KG 4PACK	2	38.44	PETCARE 2000 CARINGBAH		2229			
38	CHA0200	4506646	9310160857787	ADVANTAGE SMALL CAT 0.4KG 4PACK	1	38.44	MACKAY CITY PET BARN		4740			
39	CHA0200	4506646	9310160857787	ADVANTAGE SMALL CAT 0.4KG 4PACK	1	38.44	PET O BEVERLY HILLS		2209			
40	CHA0200	4506646	9310160857787	ADVANTAGE SMALL CAT 0.4KG 4PACK	1	38.44	PET O ROCKDALE		2216			
41	CHA0200	4506646	9310160857787	ADVANTAGE SMALL CAT 0.4KG 4PACK	4	38.44	SOUTHWEST STOCKFEEDS		2650			
42	CHA0200	4506646	9310160857787	ADVANTAGE SMALL CAT 0.4KG 4PACK	1	38.44	QUEENSLAND MAIDENVALE MADDISON VET SURV		9944			

Sheet1

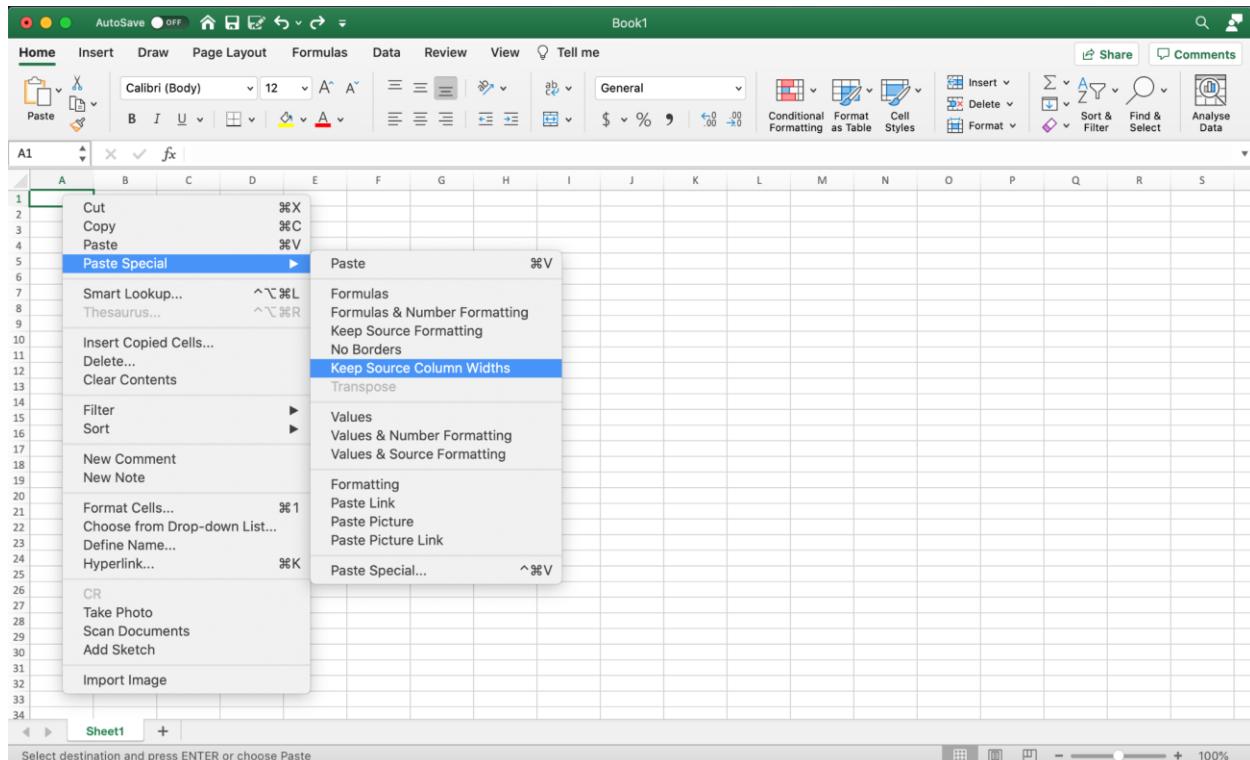
Group 1 Training Manual

Step 5: Left click the A1 cell of one of the data sets and press CTRL+A (CMD+A for mac) to select the entire table, then press right click and select Copy to copy the table.

Stockcode	Suppliercode	Barcode	Description	Quantity	Price	Acc Name	Postcode	I	J	
7265 DHS0595	10022144	5414736046952	SIMPARICA TRIO 40.1-60KG 6PACK	1	88.88	MACLEAY PETS PLUS	4184			
7266 DHS0595	10022144	5414736046952	SIMPARICA TRIO 40.1-60KG 6PACK	2	88.88	PETCARE 2000 CASTLE HILL	2154			
7267 DHS0595	10022144		Cut	1	88.88	POPLAR PETFOOD & PRODUCE	2526			
7268 DHS0595	10022144			1	88.88	PET CAFE WOODVILLE WEST	5011			
7269 DHS0595	10022144			2	88.88	WILLOWBROOK PET YARD	2746			
7270 DHS0595	10022144			3	88.88	CAMPBELLTOWN PET AQUARIUM	2560			
7271 DHS0595	10022144			-20	88.88	PETS OF SANDGATE	4017			
7272 DHS0595	10022144			2	88.88	HIGHLANDS PETS & PRODUCE	4720			
7273 DHS0595	10022144			4	88.88	RSPCA QLD	4076			
7274 DHS0595	10022144			20	88.88	PETS OF SANDGATE	4017			
7275 DHS0595	10022144			2	88.88	MACLEAY PETS PLUS	4184			
7276 DHT0220	1956075			1	7.92	PETSTOCK PORT MACQUARIE	2444			
7277 DHT0220	1956075			1	7.92	PETSTOCK MONA VALE	2103			
7278 DHT0220	1956075			2	7.92	WATERFORD RURAL CENTRE BARN	4133			
7279 DHT0220	1956075			1	7.92	PETSTOCK BATEMANS BAY	2536			
7280 DHT0220	1956075			2	7.92	PETSTOCK ADAMSTOWN	2239			
7281 DHT0220	1956075			3	7.92	PETSTOCK MOUNT COAH	2079			
7282 DHT0220	1956075			1	7.92	PETSTOCK PORT MACQUARIE	2444			
7283 DHT0220	1956075			4	7.92	COASTAL RURAL TRADERS	2258			
7284 DHT0220	1956075			2	7.92	PETSTOCK PORT MACQUARIE	2444			
7285 DHT0220	1956075			3	7.92	WAUCHOPE RURAL CENTRE	2446			
7286 DHT0220	1956075			4	7.92	PETSTOCK ERINA	2250			
7287 HHC1000	HHC1000			0	39.4	WOONONA PETFOOD & PRODUCE PTY LTD	2518			
7288 HHC1000	HHC1000			0	39.4	WOONONA PETFOOD & PRODUCE PTY LTD	2518			
7289 HHC1000	HHC1000			0	39.4	WOONONA PETFOOD & PRODUCE PTY LTD	2518			
7290 HHC1000	HHC1000			0	39.4	WOONONA PETFOOD & PRODUCE PTY LTD	2518			
7291 HHE0030	129866			IRMER 6.42G	20	9.54	PULLENS PRODUCE	2460		
7292 HHE0030	129866			IRMER 6.42G	-10	9.54	PULLENS PRODUCE	2460		
7293 HHE0030	129866			IRMER 6.42G	15	9.54	ROCHDALE PRODUCE POOL & GARDEN SUPPLIES	4127		
7294 HHE0030	129866			IRMER 6.42G	9	9.54	COMPLETE STEEL AND RURAL	2850		
7295 HHE0030	129866			IRMER 6.42G	5	9.54	WILLIAMS PET PRODUCE	2318		
7296 HHE0030	129866			IRMER 6.42G	0	9.54	MICROBERTS & SONS PTY LTD	3840		
7297 HHE0040	130233			EWORMER 15G	10	10.6	MICROBERTS & SONS PTY LTD	3840		
7298 HHE0040	130233			EWORMER 15G	5	10.6	WILLIAMS PET PRODUCE	2318		
7299 HHE0040	130233			EWORMER 15G	6	10.6	WILLIAMS PET PRODUCE	2546		
7300 HHE0040	130233			EWORMER 15G	20	10.6	WAUCHOPE RURAL CENTRE	2446		
7301 HHE0040	130233			EWORMER 15G	2	10.6	GRENfell PHARMACY	2810		
7302 MHR0020	NONE			CR	4	25.15	TOP PRODUCE & SADDLERY	4680		
7303 MHR0020	NONE			Take Photo	4	25.15	WOONONA PETFOOD & PRODUCE PTY LTD	2518		
7304 MHR0020	NONE			Scan Documents	4	25.15	HEATHCOTE RURAL MERCHANDISE	3523		
7305 MHR0020	NONE			Add Sketch	4	25.15	TOP PRODUCE & SADDLERY	4680		

Group 1 Training Manual

Step 6: Create a new spreadsheet. Then right click the A1 cell, hover on paste special, then select Keep Source Column Widths to put the table in the spreadsheet.



Group 1 Training Manual

Book1

The screenshot shows a Microsoft Excel spreadsheet titled "Book1". The ribbon menu is visible at the top, with the "Home" tab selected. The main content is a table with the following columns: Stockcode, Suppliercode, Barcode, Description, Quantity, Price, Acc Name, and Postcode. The table contains approximately 34 rows of data, mostly for "ADVANTAGE SMALL CAT 0-4KG 1PK" items from various suppliers like TAREE PET BARN, PETSTOCK TRARALGON, OAKY RURAL SUPPLIES, etc., with prices ranging from 11.97 to 38.44. The "Stockcode" column has a dropdown arrow, and the "Postcode" column has a dropdown arrow.

	A	B	C	D	E	F	G	H	I	J
1	Stockcode	Suppliercode	Barcode	Description	Quantity	Price	Acc Name	Postcode		
2	CHAO010	84266825	9310160821720	ADVANTAGE SMALL CAT 0-4KG 1PK	1	11.97	TAREE PET BARN	2430		
3	CHAO010	84266825	9310160821720	ADVANTAGE SMALL CAT 0-4KG 1PK	8	11.97	PETSTOCK TRARALGON	3844		
4	CHAO010	84266825	9310160821720	ADVANTAGE SMALL CAT 0-4KG 1PK	1	11.97	OAKY RURAL SUPPLIES	4401		
5	CHAO010	84266825	9310160821720	ADVANTAGE SMALL CAT 0-4KG 1PK	1	11.97	PETSTOCK ROSEBUD	3940		
6	CHAO010	84266825	9310160821720	ADVANTAGE SMALL CAT 0-4KG 1PK	2	11.97	PETSTOCK PORT LINCOLN	5606		
7	CHAO010	84266825	9310160821720	ADVANTAGE SMALL CAT 0-4KG 1PK	2	11.97	PETSTOCK CRANBOURNE	3977		
8	CHAO010	84266825	9310160821720	ADVANTAGE SMALL CAT 0-4KG 1PK	2	11.97	ALDINGA ONE STOP PET SHOP & AQUARIUMS	5173		
9	CHAO010	84266825	9310160821720	ADVANTAGE SMALL CAT 0-4KG 1PK	1	11.97	PETSTOCK ROSEBUD	3940		
10	CHAO010	84266825	9310160821720	ADVANTAGE SMALL CAT 0-4KG 1PK	1	11.97	PETSTOCK PORT LINCOLN	5606		
11	CHAO010	84266825	9310160821720	ADVANTAGE SMALL CAT 0-4KG 1PK	1	11.97	PETSTOCK ORMOND	3204		
12	CHAO010	84266825	9310160821720	ADVANTAGE SMALL CAT 0-4KG 1PK	1	11.97	PETSTOCK ORMOND	3204		
13	CHAO010	84266825	9310160821720	ADVANTAGE SMALL CAT 0-4KG 1PK	1	11.97	KELLYVILLE PETS	2155		
14	CHAO010	84266825	9310160821720	ADVANTAGE SMALL CAT 0-4KG 1PK	1	11.97	TAREE PET BARN	2430		
15	CHAO020	4506646	9310160657787	ADVANTAGE SMALL CAT 0-4KG 4PACK	2	38.44	PET O ANANDALE	2038		
16	CHAO020	4506646	9310160657787	ADVANTAGE SMALL CAT 0-4KG 4PACK	1	38.44	PETSTOCK TRARALGON	3844		
17	CHAO020	4506646	9310160657787	ADVANTAGE SMALL CAT 0-4KG 4PACK	1	38.44	PET O HORNSBY	2077		
18	CHAO020	4506646	9310160657787	ADVANTAGE SMALL CAT 0-4KG 4PACK	1	38.44	EDGECLIFF PETS (THE CATS WHISKERS T/AS)	2027		
19	CHAO020	4506646	9310160657787	ADVANTAGE SMALL CAT 0-4KG 4PACK	1	38.44	BEACHMERE DISCOUNT DRUG STORE	4510		
20	CHAO020	4506646	9310160657787	ADVANTAGE SMALL CAT 0-4KG 4PACK	2	38.44	PET O ROCKDALE	2216		
21	CHAO020	4506646	9310160657787	ADVANTAGE SMALL CAT 0-4KG 4PACK	1	38.44	PET O PET STORE	2232		
22	CHAO020	4506646	9310160657787	ADVANTAGE SMALL CAT 0-4KG 4PACK	1	38.44	PET O ASHFIELD	2131		
23	CHAO020	4506646	9310160657787	ADVANTAGE SMALL CAT 0-4KG 4PACK	1	38.44	PET O HORNSBY	2077		
24	CHAO020	4506646	9310160657787	ADVANTAGE SMALL CAT 0-4KG 4PACK	2	38.44	PETCARE 2000 YAGOONA	2199		
25	CHAO020	4506646	9310160657787	ADVANTAGE SMALL CAT 0-4KG 4PACK	1	38.44	CONDELL PARK PRODUCE	2200		
26	CHAO020	4506646	9310160657787	ADVANTAGE SMALL CAT 0-4KG 4PACK	1	38.44	PET O ASHFIELD	2131		
27	CHAO020	4506646	9310160657787	ADVANTAGE SMALL CAT 0-4KG 4PACK	2	38.44	IPSWICH PET & AQUARIUM	4305		
28	CHAO020	4506646	9310160657787	ADVANTAGE SMALL CAT 0-4KG 4PACK	1	38.44	PET O ALEXANDRIA	2015		
29	CHAO020	4506646	9310160657787	ADVANTAGE SMALL CAT 0-4KG 4PACK	1	38.44	PET O NORTHEMHEAD	2152		
30	CHAO020	4506646	9310160657787	ADVANTAGE SMALL CAT 0-4KG 4PACK	1	38.44	PET O ANANDALE	2038		
31	CHAO020	4506646	9310160657787	ADVANTAGE SMALL CAT 0-4KG 4PACK	2	38.44	ALDINGA ONE STOP PET SHOP & AQUARIUMS	5173		
32	CHAO020	4506646	9310160657787	ADVANTAGE SMALL CAT 0-4KG 4PACK	1	38.44	KELLYVILLE PETS	2155		
33	CHAO020	4506646	9310160657787	ADVANTAGE SMALL CAT 0-4KG 4PACK	1	38.44	PET O ANANDALE	2038		
34	CHAO020	4506646	9310160657787	ADVANTAGE SMALL CAT 0-4KG 4PACK	1	38.44	ANIMAL HOUSE BALLINA	2478		

Group 1 Training Manual

Step 7: On the other data set(s), left click the A2 cell of one of the data sets and press CTRL+A (CMD+A for mac) to select the entire table without the column headings, then right click and select Copy to copy the table.

	Stockcode	Suppliercode	Description	Quantity	Price	Acc Name	Postcode	I	J
7869	DHT0220	1986075	850646000038 TICK TWISTER	6	7.92	PETSTOCK MONA VALE	2103		
7970	DHT0220	1986075	850646000038 TICK TWISTER	2	7.92	PET QUARTERS GLENDALE	2285		
7971	DHT0220	Cut	% X MISTER	5	7.92	PETSTOCK BELROSE	2085		
7972	DHT0220		% X MISTER	1	7.92	PITTWATER PETS	2102		
7973	DHT0220		% C MISTER	2	7.92	GORDON BARRY & CO PTY LTD	2380		
7974	DHT0220		% V MISTER	3	7.92	MACLEAN DISCOUNT PHARMACY	2463		
7975	DHT0220		% MISTER	2	7.92	KEMPSEY FRIENDLIES	2440		
7976	DHT0220		> MISTER	1	7.92	PETSTOCK ADAMSTOWN	2289		
7977	DHT0220		% MISTER	1	7.92	WILSONS COUNTY PHARMACY	2160		
7978	DHT0220		% C MISTER	1	7.92	PETSTOCK ADAMSTOWN	2289		
7979	DHT0220		% V MISTER	1	7.92	PITTWATER PETS	2102		
7980	DHT0220		% C R MISTER	1	7.92	PETSTOCK BALLINA	2478		
7981	DHT0220		MISTER	1	7.92	PETSTOCK PORT MACQUARIE	2444		
7982	DHT0220		MISTER	10	7.92	PETSTOCK BALGOWLAH	2093		
7983	DHT0220		MISTER	1	7.92	PETSTOCK BATEMANS BAY	2536		
7984	HHC1000		► ITA 100 INJECTABLE 100ML	6	39.4	AUSTACK SUNSHINE COAST PTY LTD	4551		
7985	HHC1000		► ITA 100 INJECTABLE 100ML	12	39.4	BERNIES PRODUCE ADVANCECITY	4211		
7986	HHC1000		► ITA 100 INJECTABLE 100ML	30	39.4	ANIMALCARE	2756		
7987	HHC1000		► ITA 100 INJECTABLE 100ML	1	39.4	ANIMALCARE (CAMBLA ELECTRICAL SERVICES)	2760		
7988	HHC1000		► ITA 100 INJECTABLE 100ML	30	39.4	ANIMALCARE	2756		
7989	HHC1000		► ITA 100 INJECTABLE 100ML	1	39.4	WOONONA PETFOOD & PRODUCE PTY LTD	2518		
7990	HHC1000		► ITA 100 INJECTABLE 100ML	0	39.4	WOONONA PETFOOD & PRODUCE PTY LTD	2518		
7991	HHC1000		► ITA 100 INJECTABLE 100ML	4	39.4	TOP PRODUCE & SADDLERY	4680		
7992	HHC1000		► ITA 100 INJECTABLE 100ML	2	39.4	TAREE PRODUCE	2430		
7993	HHC1000		► ITA 100 INJECTABLE 100ML	2	39.4	ANIMALCARE	2756		
7994	HHC1000		► ITA 100 INJECTABLE 100ML	0	39.4	WOONONA PETFOOD & PRODUCE PTY LTD	2518		
7995	HHC1000		► ITA 100 INJECTABLE 100ML	0	39.4	WOONONA PETFOOD & PRODUCE PTY LTD	2518		
7996	HHE030	New Comment	EC PASTE HORSE WORMER 6.42G	1	9.54	PULLENS PRODUCE	2460		
7997	HHE030	New Note	EC PASTE HORSE WORMER 6.42G	0	9.54	KIRKWOOD FODDER AND FEED	4472		
7998	HHE030		EC PASTE HORSE WORMER 6.42G	0	9.54	INVERELL BARBERS	2360		
7999	HHE030		EC PASTE HORSE WORMER 6.42G	1	9.54	THE DOG HOUSE KEMPSEY	2440		
8000	HHE030		EC PASTE HORSE WORMER 6.42G	6	9.64	COMPLETE STEEL AND RURAL	2850		
8001	HHE040	Format Cells...	EC PLUS TAPE HORSE WORMER 15G	8	10.6	KIRKWOOD PRODUCE RUTHERFORD	2320		
8002	HHE040	Choose from Drop-down List...	EC PLUS TAPE HORSE WORMER 15G	4	10.6	KINCUMBER PRODUCE AND PET SUPPLIES	2251		
8003	HHE040	Hyperlink...	EC PLUS TAPE HORSE WORMER 15G	10	10.6	EQUINE PERFORMANCE	2765		
8004	HHE040		EC PLUS TAPE HORSE WORMER 15G	3	10.6	GRENfell PHARMACY	2810		
8005	HHE040		EC PLUS TAPE HORSE WORMER 15G	6	10.6	BARCOO FODDER AND FEED	4472		
8006	HHE040		EC PLUS TAPE HORSE WORMER 15G	24	10.6	JUDDS HORSE SUPPLIES	2663		
8007	HMR020	CR	IN PASTE 500G	4	25.15	GUNDAH PET PLACE	2380		
8008	HMR020	Take Photo	IN PASTE 500G	4	25.15	GOONOGOROO PET FOOD & PRODUCE PTY LTD	2518		
8009	HMR020	Scan Documents	IN PASTE 500G	4	25.15	HEATHCOTE RURAL MERCHANDISE	3523		
		Add Sketch							
		Import Image							

Step 8: On the new spreadsheet, press CTRL+ any arrow key to navigate to the end of the spreadsheet. Then right click the cell in column A under the data and press paste.

Group 1 Training Manual

The screenshot shows a Microsoft Excel spreadsheet titled "Group 1 Training Manual". The data is organized into columns A through J. Column A contains row numbers from 7987 to 8010. Columns B, C, D, E, F, and G contain product details and quantities. Columns H, I, and J contain store names and prices. A context menu is open over row 8009, listing options like Paste, Paste Special, Smart Lookup..., Thesaurus..., Insert Copied Cells..., Delete..., Clear Contents, Filter, Sort, New Comment, and New Note.

	A	B	C	D	E	F	G	H	I	J
7987	HHC1000	HHC1000	9310160820518	COFORTA 100 INJECTABLE 100ML	1	39.4	CAMBLA KENNELS (CAMBLA ELECTRICAL SERVICES)	4660		
7988	HHC1000	HHC1000	9310160820518	COFORTA 100 INJECTABLE 100ML	30	39.4	ANIMALCARE	2756		
7989	HHC1000	HHC1000	9310160820518	COFORTA 100 INJECTABLE 100ML	1	39.4	WOONONA PETFOOD & PRODUCE PTY LTD	2518		
7990	HHC1000	HHC1000	9310160820518	COFORTA 100 INJECTABLE 100ML	0	39.4	WOONONA PETFOOD & PRODUCE PTY LTD	2518		
7991	HHC1000	HHC1000	9310160820518	COFORTA 100 INJECTABLE 100ML	4	39.4	TOP PRODUCE & SADDLERY	4680		
7992	Cut			COFORTA 100 INJECTABLE 100ML	2	39.4	TAREE PRODUCE	2430		
7993	Copy			COFORTA 100 INJECTABLE 100ML	2	39.4	ANIMALCARE	2756		
7994	Paste			COFORTA 100 INJECTABLE 100ML	0	39.4	WOONONA PETFOOD & PRODUCE PTY LTD	2518		
7995	Paste Special			COFORTA 100 INJECTABLE 100ML	0	39.4	WOONONA PETFOOD & PRODUCE PTY LTD	2518		
7996	Smart Lookup...			EQUIMEC PASTE HORSE WORMER 6.42G	10	9.54	PULLENS PRODUCE	2460		
7997	Thesaurus...			EQUIMEC PASTE HORSE WORMER 6.42G	0	9.54	BARCOO FODDER AND FEED	4472		
7998				EQUIMEC PASTE HORSE WORMER 6.42G	0	9.54	INVERELL BARNYARD	2360		
7999				EQUIMEC PASTE HORSE WORMER 6.42G	1	9.54	THE DOG HOUSE KEMPSEY	2440		
8000				EQUIMEC PASTE HORSE WORMER 6.42G	6	9.54	COMPLETE STEEL AND RURAL	2850		
8001				EQUIMEC PLUS TAPE HORSE WORMER 15G	8	10.6	KIRKWOOD PRODUCE RUTHERFORD	2320		
8002				EQUIMEC PLUS TAPE HORSE WORMER 15G	4	10.6	KINCUMBER PRODUCE AND PET SUPPLIES	2251		
8003				EQUIMEC PLUS TAPE HORSE WORMER 15G	10	10.6	EQUINE PERFORMANCE	2765		
8004				EQUIMEC PLUS TAPE HORSE WORMER 15G	3	10.6	GRENfell PHARMACY	2810		
8005				EQUIMEC PLUS TAPE HORSE WORMER 15G	6	10.6	BARCOO FODDER AND FEED	4472		
8006				EQUIMEC PLUS TAPE HORSE WORMER 15G	24	10.6	JUDDS HORSE SUPPLIES	2663		
8007				RACUMIN PASTE 500G	4	25.15	GUNNEDAH PET PLACE	2380		
8008				RACUMIN PASTE 500G	4	25.15	WOONONA PETFOOD & PRODUCE PTY LTD	2518		
8009				RACUMIN PASTE 500G	4	25.15	HEATHCOTE RURAL MERCHANDISE	3523		
8010										
8011										
8012										
8013										
8014										
8015										
8016										
8017										
8018										
8019										
8020										

Repeat steps 7 and 8 if more than two data sets are to be combined.

Group 1 Training Manual

Windows Installation Guide for Anaconda

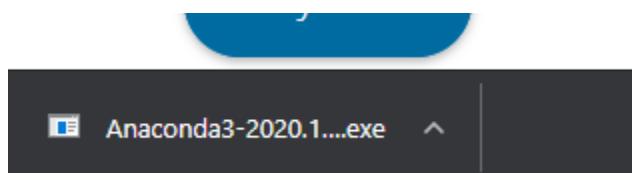
In order to install Jupyter, another base line program called ‘Anaconda’. First, ensure that you have the following minimum system requirements needed for Anaconda. The requirements are as follows:

- Operating System: Windows 8 or newer
- System Architecture: Windows 64bit x86 OR 32bit x86
- Minimum 5 GB Disk Space

After verifying system requirements, head to anaconda.com/products/individual on a search engine. After visiting the website, scroll down to the bottom, and click to download either 64bit or 32bit installer, based on your system specifications.

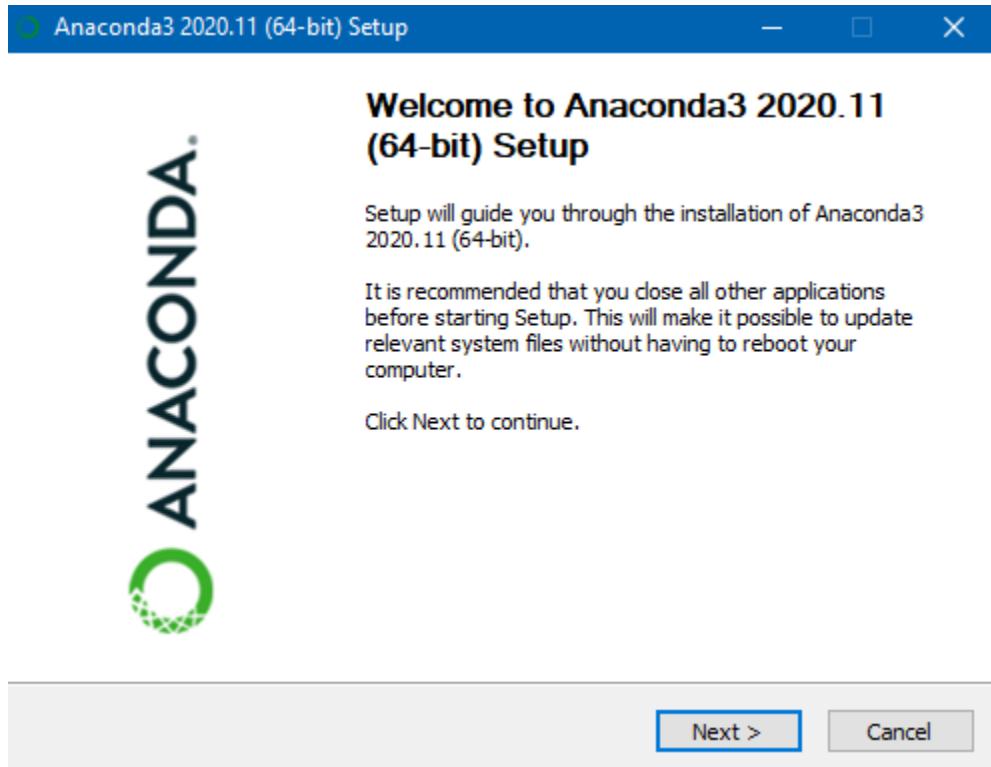


After clicking the specified installer, the program will continue to download at the bottom tab of the search engine, in this instance the user has google chrome. Continue to Wait the specified time, and then click on the Anaconda.exe file.



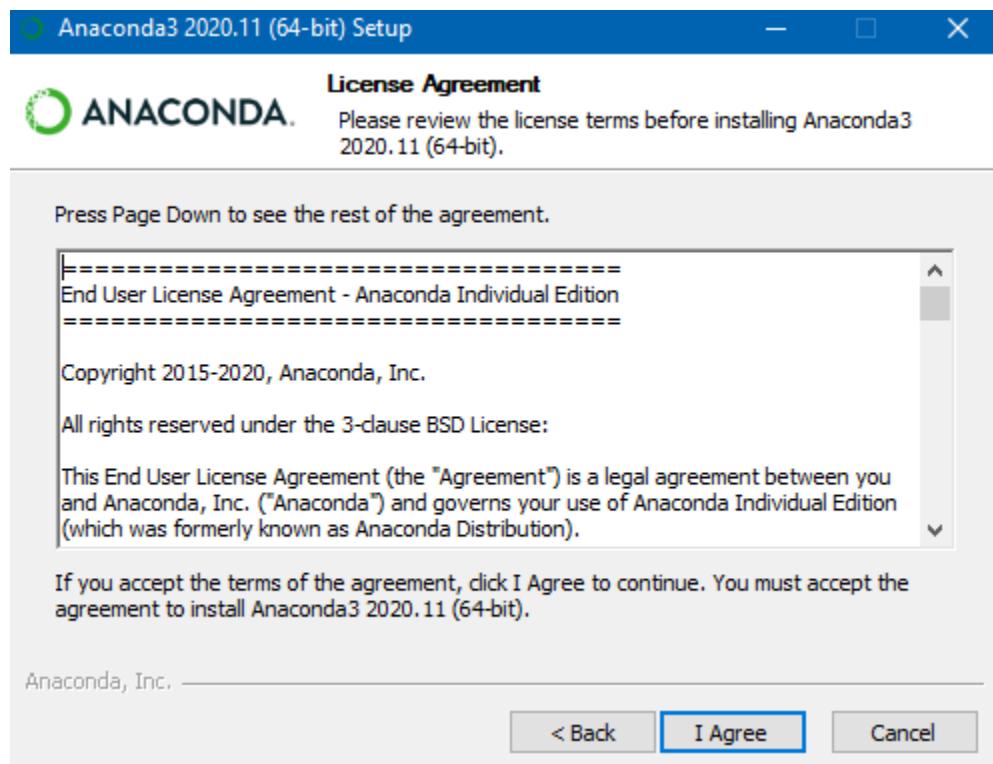
Group 1 Training Manual

You will then see this screen.



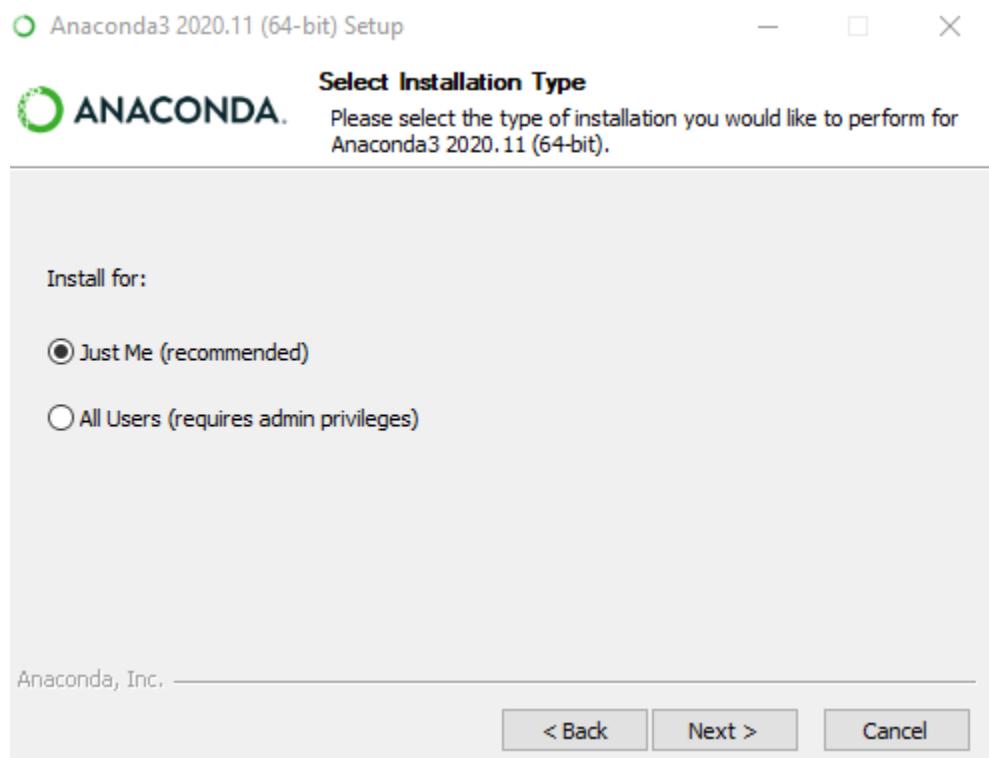
Click “Next >” which will then take you to a license agreement. Read the agreement and accept the terms and conditions by clicking “I Agree”

Group 1 Training Manual

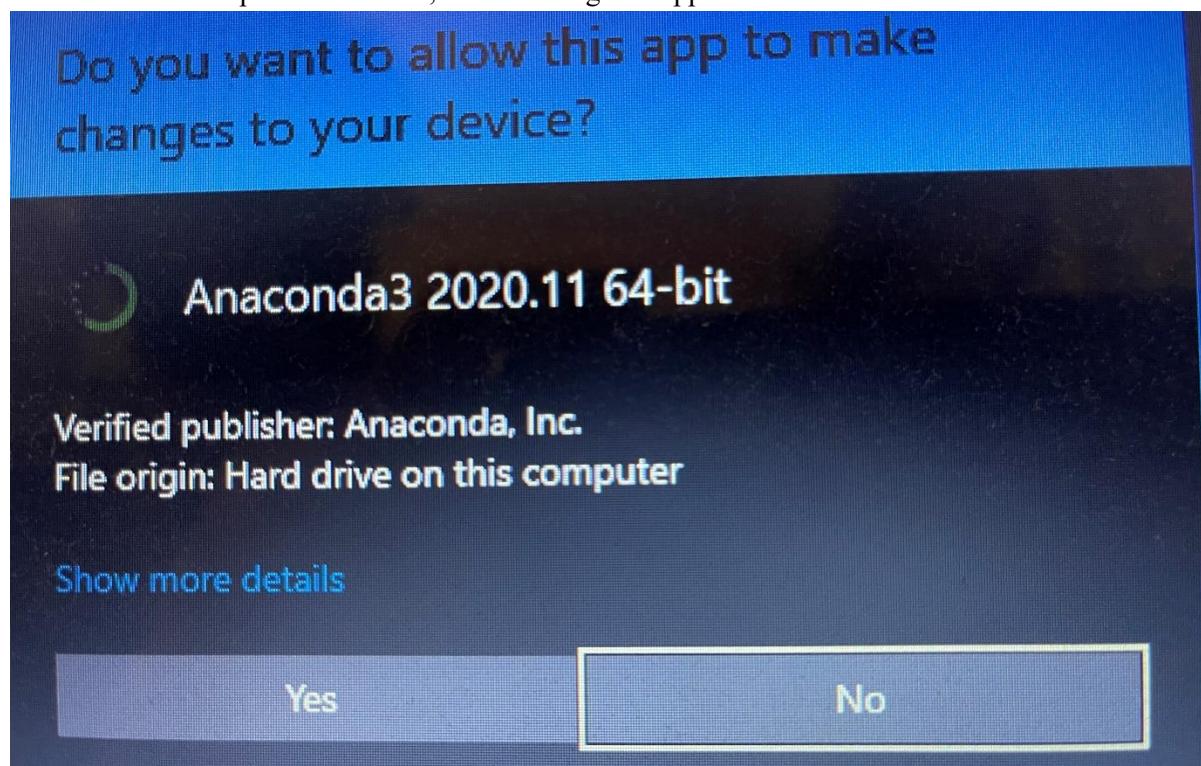


The user will be then greeted by the following screen. If the PC/Computer in use only has one user, click the “Just Me” option. Otherwise, Click the “All Users” Option.

Group 1 Training Manual



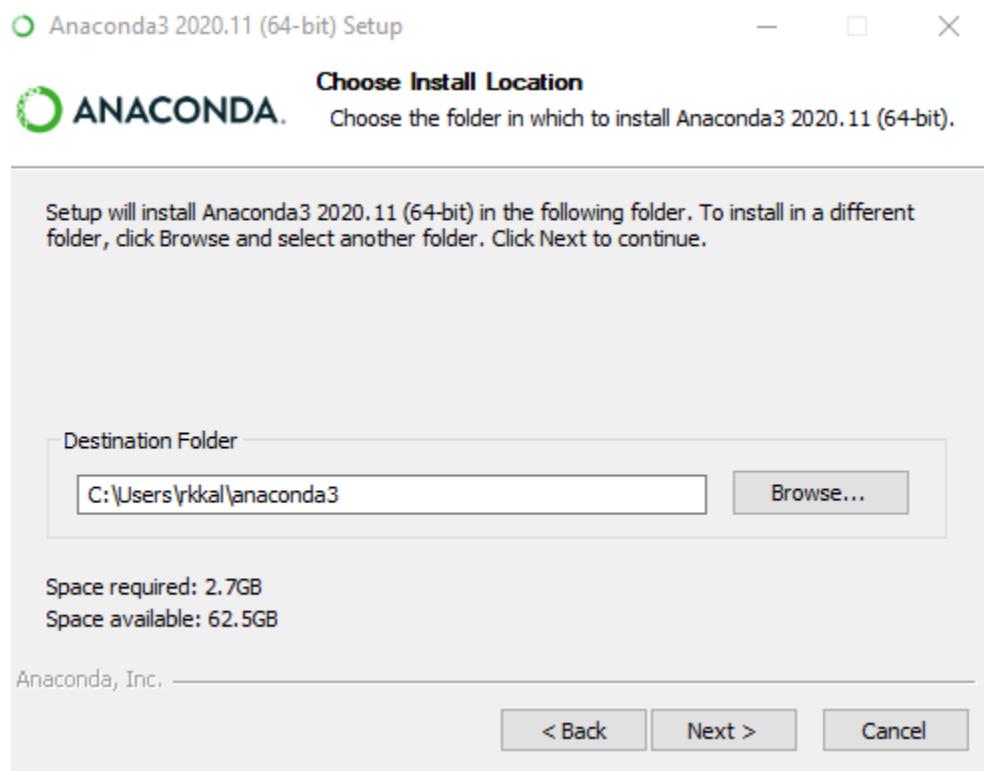
If the "All Users" option is selected, the following will appear:



Group 1 Training Manual

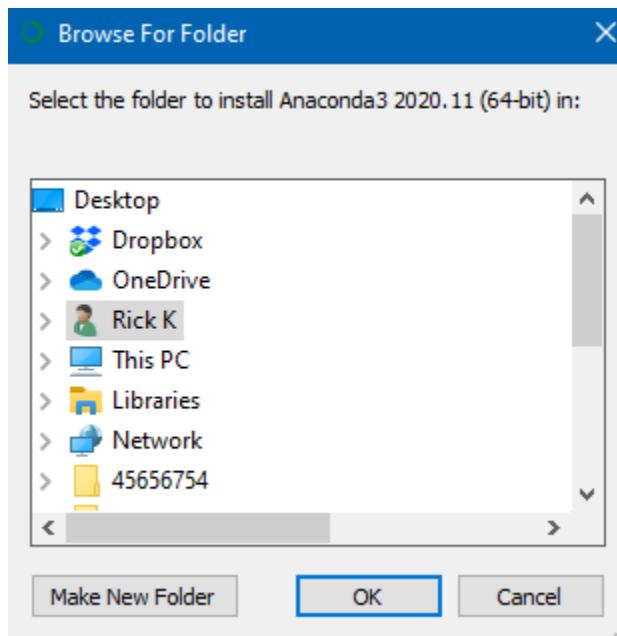
Click “Yes” and continue with the instructions below.

After clicking which type of installation the user would like, hit “Next >”. The following screen will appear:

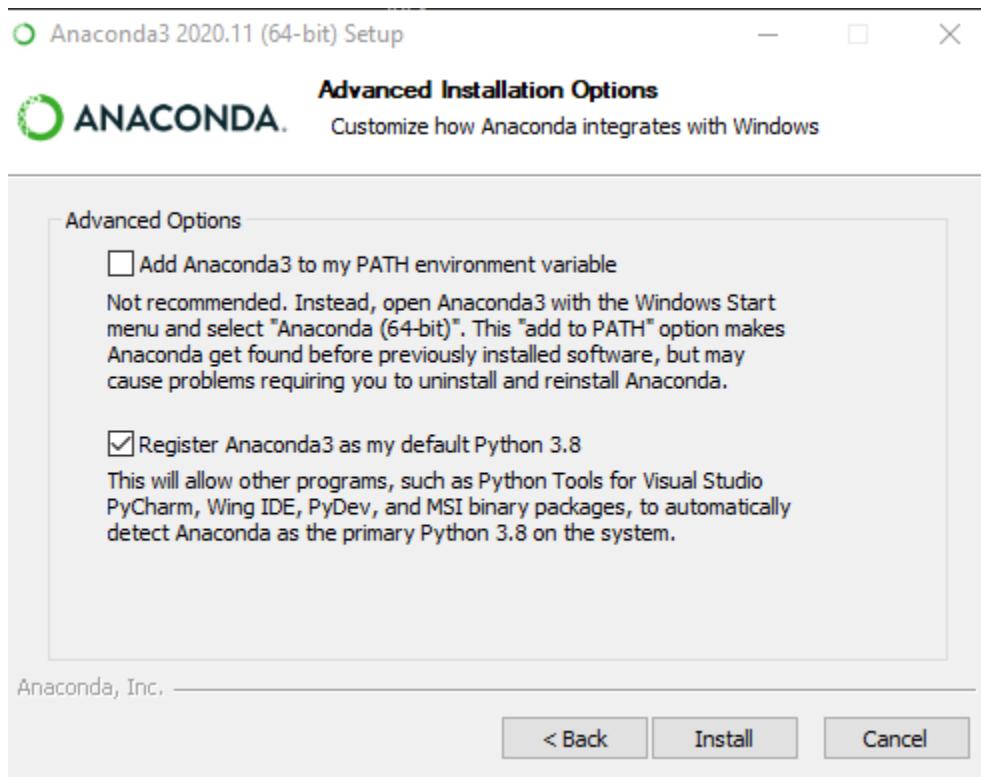


First, verify that you have the required space for the program. This space will change depending on the device you are on. Secondly, Check the destination folder to determine if the auto selected file location is plausible, or Click “Browse...” to select another file location. The following screen will appear when clicking “Browse...”

Group 1 Training Manual

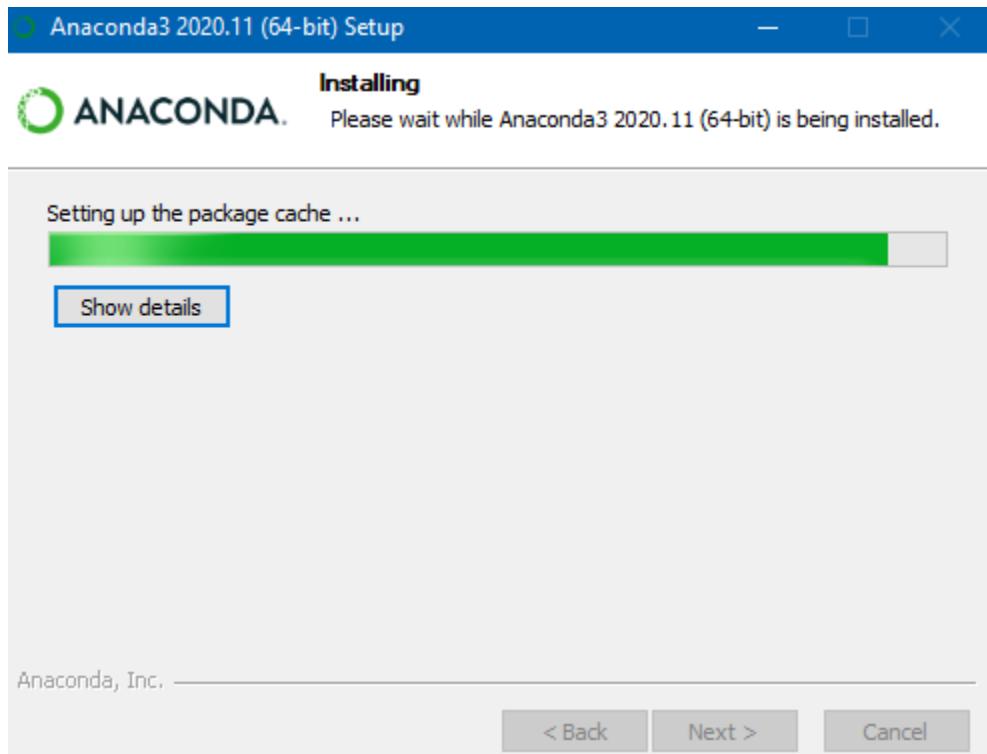


The user can then decide where Anaconda files will be stored. After making the selection, hit “OK”, and then hit “Next >” on the Previous Screen. The following screen will appear.



Group 1 Training Manual

If user wants to add Anaconda to their PATH environment variable, select the top option. If user wants to register Anaconda as the default program for other related programs, select the bottom option. If the user rejects both, deselect the second option and hit “Install”. The program will then install, followed by this screen.



When the program has finished installing, click “Next”, and then “Next” Again. The user will be greeted by this screen.

Group 1 Training Manual



Completing Anaconda3 2020.11 (64-bit) Setup

Thank you for installing Anaconda Individual Edition.

Here are some helpful tips and resources to get you started.
We recommend you bookmark these links so you can refer
back to them later.

[Anaconda Individual Edition Tutorial](#)

[Getting Started with Anaconda](#)

< Back

Finish

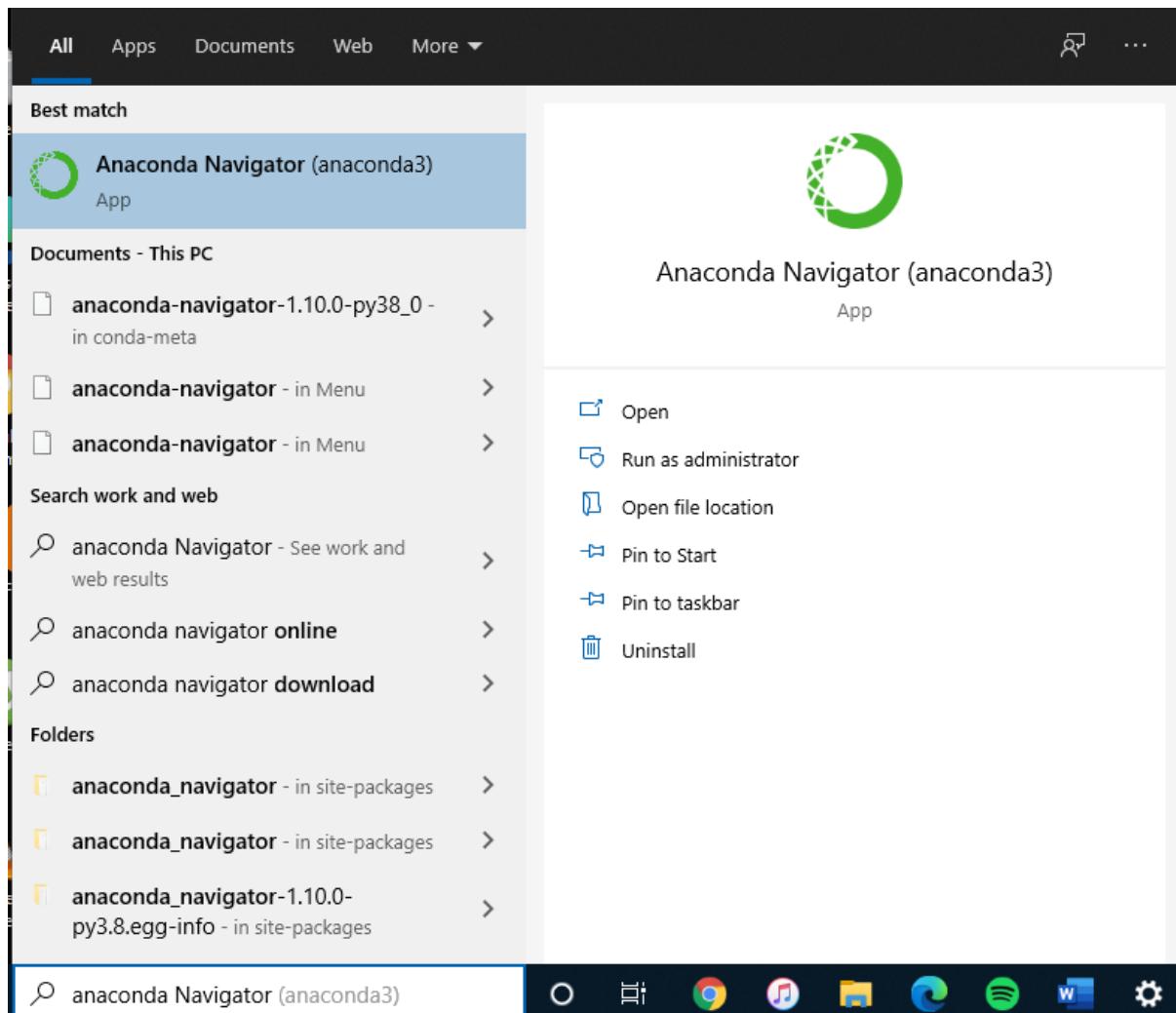
Cancel

Deselect options if familiar with Anaconda, otherwise, leave options selected and hit “Finish”.
You have successfully installed Anaconda, the base program for Jupyter.

Jupyter Access Instructions

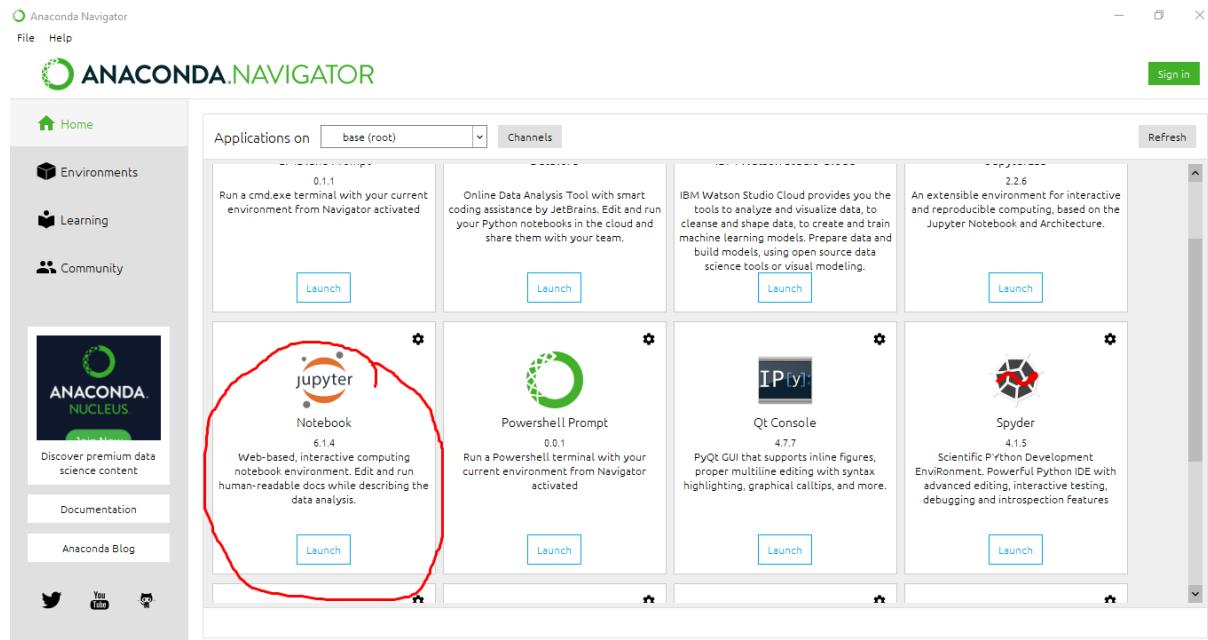
First, head to the search bar and type “Anaconda Navigator”. The program should look like the following.

Group 1 Training Manual

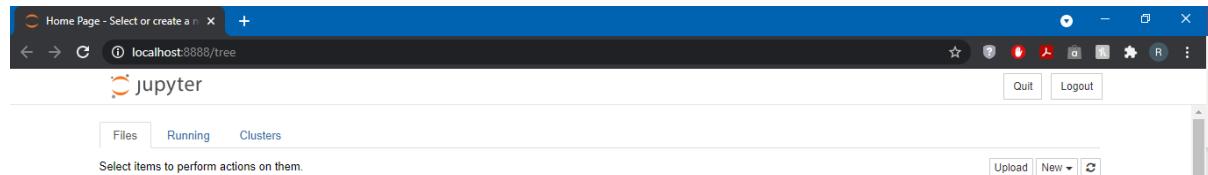


After clicking on Anaconda Navigator, and waiting some time, the following screen will appear. On the right, there is a scroll down option. Using the mouse, scroll down until you see this icon, and click “Launch”.

Group 1 Training Manual



After clicking launch, the program will load as a localhost html file on your browser. Congratulations! You have successfully accessed Jupyter.



Group 1 Training Manual

Packages

Import packages:

Import *package* as *variable*

Import features:

From *package* import *feature*

Imported packages:

- Seaborn
- Pandas
- Numpy
- Datetime
- Matplotlib.pyplot
 - imread
- Sklearn
 - Logistic regression
 - Linear regression
 - Train test split
 - RFE
 - Confusion matrix
 - Accuracy score
 - Mean Squared Error
 - R² Score
 - Polyfit
 - MLPClassifier
 - Decision tree classifier
 - K Means
 - Preprocessing
- Scipy
 - Stats
 - Personr

Dataset and Dataframe Manipulation

Data Cleaning and Manipulation

Variable = pd.read_csv(" pathname of dataset")

Imported Data:

Group 1 Training Manual

- Eastern Data Combined
- Western Combined Data
- Rename Consolidated Data Combined
- PACE Students - Market Data
- PACE Students – Wholesaler to Retail

Display column names:

Variable.columns

Aggregating Data

Drop columns:

New variable = old variable.drop(['column name', 'column name', 'column name'], axis = 1).dropna()

Note:

axis = 0: drop rows

axis = 1: drop columns

dropna(): drop null values

Merge like variables together:

New variable = old variable.groupby(['column name', 'column name']).sum().resetindex

Note:

sum(): sums the like variables together, this can be replaced with other actions

resetindex: resets the index to order by date, number and alphabetical order

Display first few rows of a table:

Variable.head()

Group 1 Training Manual

Documentation

Legend

The documentation corresponds to the MVP code written in Python. The subheadings correspond to the code and the same functions explained before are not repeated. Characteristics that can be changed to suit various purposes are depicted in *italics* (i.e. *variables*). Explanation of parameters of the code are listed under “Note:”.

Packages

```
import seaborn as sns
import pandas as pd
import numpy as np
from datetime import datetime
import matplotlib.pyplot as plt
%matplotlib inline
from sklearn.linear_model import LogisticRegression
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from sklearn.feature_selection import RFE
from sklearn.metrics import confusion_matrix, accuracy_score
from sklearn.metrics import mean_squared_error
from sklearn.metrics import mean_absolute_error
from sklearn.metrics import r2_score
from scipy import stats
from scipy.stats import pearsonr
from numpy.polynomial.polynomial import polyfit
from sklearn.neural_network import MLPClassifier
from sklearn.feature_selection import RFE
from sklearn.tree import DecisionTreeClassifier
from sklearn.tree import DecisionTreeRegressor
from sklearn.cluster import KMeans
from matplotlib.pyplot import imread
from sklearn import preprocessing

import pmdarima
import datetime
#To remove the red box errors just to make the notebook look neater
import warnings
warnings.filterwarnings("ignore")
```

Import packages:

Import package as *variable*

Import module:

From *package* import *module*

Imported packages and their modules:

- Seaborn
- Pandas
- Numpy
- Datetime
- PmdARIMA
- Warnings

Group 1 Training Manual

- Matplotlib.pyplot
 - imread
- Sklearn
 - Logistic regression
 - Linear regression
 - Train test split
 - RFE
 - Confusion matrix
 - Accuracy score
 - Mean Squared Error
 - R² Score
 - Polyfit
 - MLPClassifier
 - Decision Tree Classifier
 - Decision Tree Regressor
 - K Means
 - Preprocessing
- Scipy
 - Stats
 - Personr

Dataset and Dataframe Manipulation

Data Cleaning and Manipulation

Import Data:

```
# Import the Distribution/Sales Data into Jupyter
eastern = pd.read_csv("Data Extract/CSV/EASTERN-DATA-COMBINED.csv")
western = pd.read_csv("Data Extract/CSV/WESTERN-COMBINED-DATA.csv")
combined = pd.read_csv("Data Extract/CSV/RENAME-CONSOLIDATED-DATA-COMBINED.csv")
market = pd.read_csv("Data Extract/CSV/PACE students - market data (confidential).csv")
wholesaler_to_retail = pd.read_csv("Data Extract/CSV/PACE students - wholesaler to retail.csv",encoding='cp1252')
```

Variable = pd.read_csv(" pathname of dataset")

Imported Data:

- Eastern Data Combined
- Western Combined Data
- Rename Consolidated Data Combined
- PACE Students - Market Data
- PACE Students – Wholesaler to Retail

Display column names:

```
wholesaler_to_retail.columns
```

Variable.columns

Group 1 Training Manual

Aggregating Data

Drop columns:

```
market_drop = market.drop(['CalYr', 'MAT', 'YTD',
    'Brand', 'Product', 'ProductCode',
    'ProductionCompanion', 'Region', 'Units', 'Units ADJ', 'Doses',
    'Manufacturer ADJ', 'Brand ADJ', 'Product ADJ',
    'SubCategory3.1', 'Species ADJ', 'Doses ADJ',
    'Doses / Unit', 'Monthly Doses',
    'BI Market Only FLG'], axis=1).dropna()
```

New variable = old variable.drop(['column name', 'column name', 'column name'], axis = 1).dropna()

Note:

- axis = 0: drop rows
- axis = 1: drop columns
- dropna(): drop null values

Merge like variables together:

```
# Merge data to Period with Units sold
market_drop_unit = market_drop.groupby(['Period', 'Qtr', 'Manufacturer', 'SubCategory2', 'SubCategory3', 'Species',
    'YEAR', 'BI Business Unit']).sum().reset_index()
market_drop_unit
```

New variable = old variable.groupby(['column name', 'column name']).sum().resetindex

Note:

- sum(): sums the like variables together, this can be replaced with other actions
- resetindex: resets the index to order by date, number and alphabetical order

Display first few rows of a table:

```
wholesaler_to_retail.head()
```

Variable.head()

Data Visualisation / Exploration

Plot line graphs of each parameter (BI Business Units):

```
market_drop_unit.loc[market_drop_unit['BI Business Unit'] == 'Cattle & Sheep'].plot(x='Period', y='Value', figsize=(20,5), c = "red")
market_drop_unit.loc[market_drop_unit['BI Business Unit'] == 'Equine'].plot(x='Period', y='Value', figsize=(20,5), c = "blue")
market_drop_unit.loc[market_drop_unit['BI Business Unit'] == 'No Bus'].plot(x='Period', y='Value', figsize=(20,5), c = "green")
market_drop_unit.loc[market_drop_unit['BI Business Unit'] == 'Pets'].plot(x='Period', y='Value', figsize=(20,5), c = "black")
market_drop_unit.loc[market_drop_unit['BI Business Unit'] == 'Swine'].plot(x='Period', y='Value', figsize=(20,5), c = "orange")
```

Group 1 Training Manual

`Dataframe.loc[dataframe['column name'] == parameter].plot(x= 'variable', y='variable', figsize = (number, number), c= "colour")`

Note:

.loc: Accesses rows or columns by labels

.plot: To plot

x= : Name of variable to be plotted on x-axis

y= : Name of variable to be plotted on y-axis

figsize= : Size of the plot

c= : Colour of the line

Plot period vs value of each parameter together (BI Business Units):

```
x1 = market_drop_unit.loc[market_drop_unit['BI Business Unit'] == 'Cattle & Sheep']['Period']
y1 = market_drop_unit.loc[market_drop_unit['BI Business Unit'] == 'Cattle & Sheep']['Value']
x2 = market_drop_unit.loc[market_drop_unit['BI Business Unit'] == 'Equine']['Period']
y2 = market_drop_unit.loc[market_drop_unit['BI Business Unit'] == 'Equine']['Value']
x3 = market_drop_unit.loc[market_drop_unit['BI Business Unit'] == 'No Bus']['Period']
y3 = market_drop_unit.loc[market_drop_unit['BI Business Unit'] == 'No Bus']['Value']
x4 = market_drop_unit.loc[market_drop_unit['BI Business Unit'] == 'Pets']['Period']
y4 = market_drop_unit.loc[market_drop_unit['BI Business Unit'] == 'Pets']['Value']
x5 = market_drop_unit.loc[market_drop_unit['BI Business Unit'] == 'Swine']['Period']
y5 = market_drop_unit.loc[market_drop_unit['BI Business Unit'] == 'Swine']['Value']
x_tick = ['2018-01', '2019-01', '2020-01', '2020-12']
plt.figure(figsize=(20, 6))
plt.plot(x1, y1,label ="CS")
plt.plot(x2, y2,label ="E")
plt.plot(x3, y3,label ="NB")
plt.plot(x4, y4,label ="P")
plt.plot(x5, y5,label ="S")
plt.xticks(x_tick)
plt.legend()
```

`Variable = dataframe.loc[dataframe['column name'] == parameter]`

`Variable.plot(x variable, y variable, label = "name")`

`Variable.xticks("x axis labels")`

`Variable.legend()`

Note:

Label = : Legend label name

.xticks: x axis labels

.legend(): Include legend

Several lines can be plotted together using .plot

Group 1 Training Manual

Modelling

Creating Model Dataframe

Find and sort unique items in an array:

```
Period = market_drop['Period']
Period_Unique = Period.unique()

Qtr = market_drop['Qtr']
Qtr_Unique = Qtr.unique()

Manufacturer = market_drop['Manufacturer']
Manufacturer_Unique = Manufacturer.unique()

SubCategory2 = market_drop['SubCategory2']
SubCategory2_Unique = SubCategory2.unique()

SubCategory3 = market_drop['SubCategory3']
SubCategory3_Unique = SubCategory3.unique()

Species = market_drop['Species']
Species_Unique = Species.unique()

BI_Business_Unit = market_drop['BI Business Unit']
BI_Business_Unit_Unique = BI_Business_Unit.unique()
```

Variable.unique()

Transform labels into numerical labels:

```
le = preprocessing.LabelEncoder()

le.fit(Period_Unique)
arr1 = le.transform(Period)

le.fit(Qtr_Unique)
arr2 = le.transform(Qtr)

le.fit(Manufacturer_Unique)
arr3 = le.transform(Manufacturer)

le.fit(SubCategory2_Unique)
arr4 = le.transform(SubCategory2)

le.fit(SubCategory3_Unique)
arr5 = le.transform(SubCategory3)

le.fit(Species_Unique)
arr6 = le.transform(Species)

le.fit(BI_Business_Unit_Unique)
arr7 = le.transform(BI_Business_Unit)
```

variable = preprocessing.LabelEncoder()

variable.fit(column name)

variable.transform(column name)

Note:

.fit: Order numerical labels

.transform: Reassign labels to the data

Create a new table:

Group 1 Training Manual

```
market_drop_model = pd.DataFrame({'Period':arr1, 'Qtr':arr2, 'Manufacturer':arr3, 'SubCategory2':arr4, 'SubCategory3':arr5, 'Species':arr6, 'BI Business Unit':arr7, 'Year':market_drop['YEAR'], 'Value':market_drop['Value'].astype(int)})
```

pd.DataFrame({‘Column Name’: Data, ‘Column Name’: Data})

Create a scatterplot matrix:

```
sns.pairplot(market_drop_model)
```

sns.pairplot(dataframe)

Perform linear regression:

```
model = LinearRegression()
X = market_drop_model[['Period', 'Qtr', 'Manufacturer', 'SubCategory2', 'SubCategory3',
                      'Species', 'BI Business Unit', 'Year']]
y = market_drop_model[['Value']]
X_train, X_test = train_test_split(X, test_size=0.4, random_state=7)
y_train, y_test = train_test_split(y, test_size=0.4, random_state=7)

model.fit(X_train,y_train)
# Obtain MSE and r2 for testing data
y_test_pred = model.predict(X_test)
print("Checking testing values")
print("RMSE: ", mean_squared_error(y_test,y_test_pred,squared=False))
print("MAE: " , mean_absolute_error(y_test, y_test_pred))
print("R2: " , r2_score(y_test, y_test_pred))
```

Model = LinearRegression()

X = dataframe[[‘Column Name’, ‘Column Name’]]

y = dataframe[[‘Column Name’]]

X_train, X_test = train_test_split(X, test_size = number, random_state = number)

y_train, y_test = train_test_split(y, test_size = number, random_state = number)

Model.fit(X_train, y_train)

y_test_pred = Model.predict(X_test)

print(mean_squared_error(y_test, y_test_pred, squared = False))

print(mean_absolute_error(y_test, y_test_pred))

print(r2_score(y_test, y_test_pred))

Note:

test_size: size of testing data in a percentage

random_state: to guarantee output is the same when ran twice

mean_squared_error: Mean squared error calculator

mean_absolute_error: Mean absolute error calculator

r2_score: R^2 score calculator

This can be performed using one or more x variables

Group 1 Training Manual

KNN

```

X = market_drop_model[['Period', 'Qtr', 'Manufacturer', 'SubCategory2', 'SubCategory3',
                      'Species', 'BI Business Unit', 'Year']]
y = market_drop_model[['Value']]
X_train, X_test = train_test_split(X, test_size=0.4, random_state=7)
y_train, y_test = train_test_split(y, test_size=0.4, random_state=7)

from sklearn.neighbors import KNeighborsClassifier
K_Neighbour = KNeighborsClassifier(n_neighbors=1)

K_Neighbour.fit(X_train,y_train)
y_train_pred = K_Neighbour.predict(X_train)
y_test_pred = K_Neighbour.predict(X_test)
#print('The accuracy of the training with all the variables KNearestNeighbour model is : ', accuracy_score(y_train, y_trainpred))
#print('The accuracy of the testing with all the variables KNearestNeighbour model is : ', accuracy_score(y_test, y_testpred))

print("Checking testing values")
print("RMSE: ", mean_squared_error(y_test,y_test_pred,squared=False))
print("MAE: " , mean_absolute_error(y_test, y_test_pred))
print("R2: " , r2_score(y_test, y_test_pred))

```

```
from sklearn.neighbors import KNeighborsClassifier  
K_Neighbour = KNeighborsClassifier(n_neighbors=number)
```

```
K_Neighbour.fit(X_train,y_train)
y_train_pred = K_Neighbour.predict(X_train)
y_test_pred = K_Neighbour.predict(X_test)

print("Checking testing values")
print("RMSE: ", mean_squared_error(y_test,y_test_pred,squared=False))
print("MAE: " , mean_absolute_error(y_test, y_test_pred))
print("R2: " , r2_score(y_test, y_test_pred))
```

Note:

n_neighbors: Number of neighbours

ARIMA

Display first rows of a table:

```
market_drop.info()
```

Dataframe.info()

Import ARIMA modelling package:

Group 1 Training Manual

This also has to be performed through terminal using the commands as follows

```
(base) C:\Users\thoma>pip install pmdarima
Collecting pmdarima
  Using cached pmdarima-1.8.2-cp38-cp38-win_amd64.whl (596 kB)
Requirement already satisfied: scikit-learn>=0.22 in c:\users\thoma\anaconda3\lib\site-packages (from pmdarima) (0.23.2)

Requirement already satisfied: numpy~1.19.0 in c:\users\thoma\anaconda3\lib\site-packages (from pmdarima) (1.19.2)
Requirement already satisfied: joblib<0.11 in c:\users\thoma\anaconda3\lib\site-packages (from pmdarima) (0.17.0)
Requirement already satisfied: setuptools!=50.0.0,>=38.6.0 in c:\users\thoma\anaconda3\lib\site-packages (from pmdarima) (50.3.1.post20201107)
Requirement already satisfied: scipy>=1.3.2 in c:\users\thoma\anaconda3\lib\site-packages (from pmdarima) (1.5.2)
Requirement already satisfied: pandas>=0.19 in c:\users\thoma\anaconda3\lib\site-packages (from pmdarima) (1.1.3)
Requirement already satisfied: Cython!=0.29.18,>=0.29 in c:\users\thoma\anaconda3\lib\site-packages (from pmdarima) (0.29.1)
Requirement already satisfied: urllib3 in c:\users\thoma\anaconda3\lib\site-packages (from pmdarima) (1.25.11)
Requirement already satisfied: statsmodels!=0.12.0,>=0.11 in c:\users\thoma\anaconda3\lib\site-packages (from pmdarima) (0.12.2)
Requirement already satisfied: threadpoolctl>=2.0.0 in c:\users\thoma\anaconda3\lib\site-packages (from scikit-learn>=0.22->pmdarima) (2.1.0)
Requirement already satisfied: pytz>=2017.2 in c:\users\thoma\anaconda3\lib\site-packages (from pandas>=0.19->pmdarima) (2020.1)
Requirement already satisfied: python-dateutil>=2.7.3 in c:\users\thoma\anaconda3\lib\site-packages (from pandas>=0.19->pmdarima) (2.8.1)
Requirement already satisfied: patsy>=0.5 in c:\users\thoma\anaconda3\lib\site-packages (from statsmodels!=0.12.0,>=0.11->pmdarima) (0.5.1)
Requirement already satisfied: six>=1.5 in c:\users\thoma\anaconda3\lib\site-packages (from python-dateutil>=2.7.3->pandas>=0.19->pmdarima) (1.15.0)
Installing collected packages: pmdarima
```

Create call function:

```
def plotarima(n_periods, timeseries, automodel):
    # Forecast
    fc, confint = automodel.predict(n_periods=n_periods, return_conf_int=True)
    # Weekly index
    fc_ind = pd.date_range(timeseries.index[timeseries.shape[0]-1], periods=n_periods, freq= "W")
    # Forecast series
    fc_series = pd.Series(fc, index=fc_ind)
    # Upper and lower confidence bounds
    lower_series = pd.Series(confint[:, 0], index=fc_ind)
    upper_series = pd.Series(confint[:, 1], index=fc_ind)
    # Create plot
    plt.figure(figsize=(10, 6))
    plt.plot(timeseries)
    plt.plot(fc_series, color="colour")
    plt.xlabel("date")
    plt.ylabel(timeseries.name)
    plt.fill_between(lower_series.index,
                    lower_series,
                    upper_series,
                    color="colour",
                    alpha=number)
    plt.legend(("column names"), loc="location")
    plt.show()
```

Note:

Freq: Common types are M, W or BM. M is month, W is week and BM is business month end

Group 1 Training Manual

Confint: Confidence interval
[:, 0]: First row to last row of column 0
Loc: Location of the legend
Alpha: Confidence level

Call the ARIMA function:

```
automodel = arimamodel(market_time_series_model["Value"])
plotarima(36, market_time_series_model["Value"], automodel)
```

plotarima(*number of time periods*)

DTR

Decision Tree Regressor:

```
x = market_drop_model[['Period', 'Qtr', 'Manufacturer', 'SubCategory2', 'SubCategory3',
                      'Species', 'BI Business Unit', 'Year']]
y = market_drop_model[['Value']]
x_train, x_test = train_test_split(x, test_size=0.4, random_state=7)
y_train, y_test = train_test_split(y, test_size=0.4, random_state=7)

regressor = DecisionTreeRegressor()
regressor.fit(x_train, y_train)

y_train_pred = regressor.predict(x_train)
y_test_pred = regressor.predict(x_test)

print("Checking testing values")
print("RMSE: ", mean_squared_error(y_test,y_test_pred,squared=False))
print("MAE: " , mean_absolute_error(y_test, y_test_pred))
print("R2: " , r2_score(y_test, y_test_pred))
```

```
X = dataframe[['Column Name', 'Column Name']]
y = dataframe[['Column Name']]
X_train, X_test = train_test_split(X, test_size = number, random_state = number)
y_train, y_test = train_test_split(y, test_size = number, random_state = number)
```

Model = DecisionTreeRegressor()

Model.fit(X_train, y_train)

```
y_train_pred = Model.predict(X_train)
y_test_pred = Model.predict(X_test)
```

```
print(mean_squared_error(y_test, y_test_pred, squared = False))
print(mean_absolute_error(y_test, y_test_pred))
print(r2_score(y_test, y_test_pred))
```

Group 1 Training Manual

Logistic Regression

Split into test and train as depicted above and view the shape:

```
print("X_train shape :", X_train.shape)
print("y_train shape : ", y_train.shape)
print("X_test shape :", X_test.shape)
print("y_test shape : ", y_test.shape)
```

Print(*variable*.shape)

Perform recursive feature elimination:

```
model = LogisticRegression()
rfe = RFE(estimator=model, n_features_to_select=3, step=1)
rfe.fit(X_train, y_train)

y_pred = rfe.predict(X_test)
print("Test accuracy",3,":", accuracy_score(y_test, y_pred))
print(rfe.support_)
column_Names = X_train.columns
print(column_Names)
```

rfe = RFE(*estimator* = LogisticRegression(), *n_features_to_select* = *number*, *step* = *number*)

Variable(*X_train*, *y_train*)

```
Y_pred = rfe.predict(X_test)

print("Test accuracy",number,":", accuracy_score(y_test, y_pred))
print(rfe.support_)
column_Names = X_train.columns
print(column_Names)
```

Test the logistic regression model with different number of features:

```
# Testing a Logistic Regression model with different number of features
model = LogisticRegression()

for i in range(1,7):
    rfe = RFE(estimator=model, n_features_to_select=i, step=1)
    rfe.fit(X_train, y_train)

    y_pred = rfe.predict(X_test)
    print("Test accuracy",i,":", accuracy_score(y_test, y_pred))
    print(rfe.support_)
    column_Names = X_train.columns
    print(column_Names)
    #print("Confusion matrix",i)
    #print(confusion_matrix(y_test, y_pred))
    #print()
```

model = LogisticRegression()

```
for i in range(number,number):
    rfe = RFE(estimator=model, n_features_to_select=i, step=1)
    rfe.fit(X_train, y_train)

    y_pred = rfe.predict(X_test)
    print("Test accuracy",i,":", accuracy_score(y_test, y_pred))
    print(rfe.support_)
    column_Names = X_train.columns
    print(column_Names)
```

Group 1 Training Manual

Note:

The same code as above but a for loop is created to test the model.

Prediction

Choosing Parameters

In the Jupyter Notebook, the final section is called prediction and we are using this section to retrieve an expected/predicted output for our model.

PREDICTIONS

From the inputs, re-encode a new data set and run the prediction

Below are the "arr" listed to compare them to the actual parameter to be used

```
le.fit(Period_Unique) arr1 = le.transform(Period)

le.fit(Qtr_Unique) arr2 = le.transform(Qtr)

le.fit(Manufacturer_Unique) arr3 = le.transform(Manufacturer)

le.fit(SubCategory2_Unique) arr4 = le.transform(SubCategory2)

le.fit(SubCategory3_Unique) arr5 = le.transform(SubCategory3)

le.fit(Species_Unique) arr6 = le.transform(Species)

le.fit(BI_Business_Unit_Unique) arr7 = le.transform(BI_Business_Unit)
```

Based on the dataframes above, we will decide our encoding for the data. Additionally, this will also allow us to select a parameter we want to test.

Create Prediction parameters

```
In [39]: X = market_drop_model[['Qtr','Manufacturer']]
y = market_drop_model[['Value']]
X_train, X_test = train_test_split(X, test_size=0.4, random_state=7)
y_train, y_test = train_test_split(y, test_size=0.4, random_state=7)

from sklearn.neighbors import KNeighborsClassifier
K_Neighbour = KNeighborsClassifier(n_neighbors=1)

K_Neighbour.fit(X_train,y_train)
y_train_pred = K_Neighbour.predict(X_train)
y_test_pred = K_Neighbour.predict(X_test)
#print('The accuracy of the training with all the variables KNearestNeighbour model is : ', accuracy_score(y_train, y_trainpred))
#print('The accuracy of the testing with all the variables KNearestNeighbour model is : ', accuracy_score(y_test, y_testpred))

print("Checking testing values")
print("RMSE: ", mean_squared_error(y_test, y_test_pred,squared=False))
print("MAE: " , mean_absolute_error(y_test, y_test_pred))
print("R2: " , r2_score(y_test, y_test_pred))

Checking testing values
RMSE:  9835.652146047632
MAE: 23647.5076283637
R2: 0.9657913955424002
```

In order to select the parameters for the testing, you need to first find the encoding as done above and then choose the parameters you want in the array

```
In [40]: X_new = np.array([[12,2]])

In [41]: prediction = K_Neighbour.predict(X_new)
print("prediction: ", prediction)

prediction: [178]
```

From the image above, we can see that the model used here is the KNN model and we have created the X_new as our prediction table.

Group 1 Training Manual

After inputting the value that is used in the encoding, just run the section of code and this will give you the prediction result below.