# Lecture Note #20: Data Analysis Tools Part #2

## BUSI 201: Business Data Analysis

## Fall 2023

#### **Topic 1. Manually Importing Data**

Sometimes, you will be the one recording data in an Excel spreadsheet. But sometimes, you will be importing data from outside sources into Excel to perform data analysis. We can either manually import data ourselves, or rely on built-in tools that Excel has to offer. First, we will examine some basic manual data importing from outside sources.

Suppose you are interested secondary education attainment around the world. A quick search may lead you to a Wikipedia article titled "List of countries by secondary education attainment."<sup>1</sup> Figure 1 below is a screenshot of said webpage captured as of November 2023.

The Free Encyclopedia	Q Search Wikipedia		Sea	ch								Create account Log in
	List of countries b	y seco:	ndary educ	ation	attaiı	nmei	nt			文 <sub>A</sub> 11	anguage 🗸	
ntents [hide]	Article Talk							Read	Edit V	iew histor	ry Tools 🗸	
)	From Wikipedia, the free encyclopedi	а						_				
of countries by percent ondary education attainment for cted age groups orences	This is a list of countries by the proportion of the relevant age groups sources.											
	List of countries by perce	• Year •	dary education 3 to 5 Years above graduation age (%) <sup>[1]</sup>	Year	20-24 (%) <sup>[2]</sup> •	Year	20-29 (%) <sup>[3]</sup> •	Year	25-29	Year ¢	25–34 (%) <sup>[4]</sup> ♦	
	South Korea											
		2014	99			2015	98					
	Georgia	2014	99			2015 2013	98 95					
	+++ Georgia	2013	96	2015	95.7							
	Georgia  Japan	2013 2016	96 95	2015	95.7	2013	95					
	Georgia       Japan       Croatia	2013 2016 2013	96 95 95	2015	95.7	2013 2013	95 95			2015	82	
	Georgia     Japan     Croatia     Ukraine	2013 2016 2013 2012	96 95 95 95			2013 2013 2012	95 95 94			2015 2015	82 91	
	++ Georgia Japan Croatia Ukraine Sweden	2013 2016 2013 2012	96 95 95 95	2015	87.3	2013 2013 2012	95 95 94					
	++     Georgia       Japan       Croatia       Ukraine       Sweden       Ireland	2013 2016 2013 2012 2013	96 95 95 95 92	2015	87.3	2013 2013 2012 2012	95 95 94 94					
	Georgia     Japan     Croatia     Ukraine     Sweden     Ireland     Armenia	2013 2016 2013 2012 2013 2013 2010	96 95 95 92 92 93	2015 2015	87.3 92.7	2013 2013 2012 2012 2012 2010	95 95 94 94 94			2015	91	
	Georgia     Japan     Croatia     Ukraine     Sweden     Ireland     Armenia     United Kingdom	2013 2016 2013 2012 2013 2013 2010 2013	96 95 95 92 92 93 94	2015 2015	87.3 92.7	2013 2013 2012 2012 2012 2010 2013	95 95 94 94 94 92 92			2015	91	

Figure 1: Wikipedia Article

One way to import this data into Excel is to simply copy and paste the entire table. You can copy the data in the table by left clicking and dragging to select the table, and then right clicking the selected table, and selecting copy.

<sup>&</sup>lt;sup>1</sup>https://en.wikipedia.org/wiki/List\_of\_countries\_by\_secondary\_education\_attainment

You can choose two options when pasting the table data into Excel. You can choose to Keep Source Formatting, or Match Destination Formatting as shown in Figure 2. We will primarily be using the latter, as the source formatting is not necessarily well translated over to Excel.

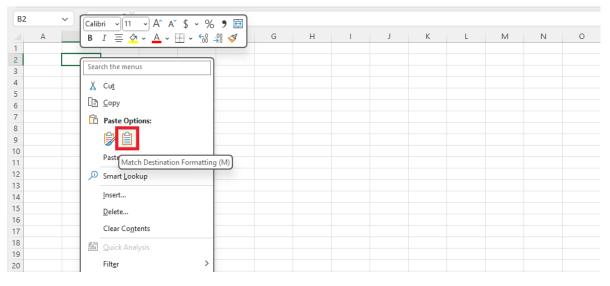


Figure 2: Pasting Options

Pasting the table we copied earlier while matching destination formatting, we can import the table as shown in Figure 3. Now that we have the table in Excel, we can use the tools that we have at our disposal to "clean" the data. Remove redundant rows and columns, sorting data by educational attainment, color-coding the table using conditional formatting, creating charts to visualize data, etc.

	А	В	С	D	E	F	G	Н	1	J	К	L	М	N	0
1															
2															
3		Country	Year	3 to 5 Yea	Year	20-24	Year	20-29	Year	25-29	Year	25-34			
4				graduatio	n age	(%)[2]		(%) [3]		(%)		(%)[4]			
5				(%)[1]											
6		South Ko	2014	99			2015	98							
7		Georgia	2013	96			2013	95							
8		Japan	2016	95											
9		Croatia	2013	95	2015	95.7	2013	95							
10		Ukraine	2012	95			2012	94							
11		Sweden	2013	92	2015	87.3	2012	94			2015	82			
12		Ireland			2015	92.7					2015	91			
13		Armenia	2010	93			2010	92							
14		United Ki	2013	94	2015	85.7	2013	92			2015	85			
15		Kazakhsta	2010	93			2010	92							
16		Poland	2013	83	2015	90.8	2013	92			2015	94			
17		United St	2010	92			2013	91			2015	90			
18		Canada	2010	86			2010	91			2015	93			
19		Greece	2013	92	2015	89.6	2013	91			2015	84			
20		Slovakia	2013	93	2015	91.3	2013	90			2015	93			

Figure 3: Imported Table

This is a rather straightforward example of importing data. The original source material was already formatted as a table, and the importing process required little customization. Now, let us examine a case where the data requires a bit more work

#### **CSV:** Comma Separated Values

In some cases, you will encounter files in the form of pdfs or txt files. One such example can be found by downloading the BUSI201-LEC20-txt file. This file lists the top 20 movies of all time based on IMDB review scores as of November 2023.

```
BUSI201-LEC20-txt
                                 ×
                                       ^+
File
      Edit
             View
Title, Year, Length, IMDB Rating
The Shawshank Redemption, 1994, 142, 9.3
The Godfather, 1972, 175, 9.2
The Dark Knight, 2008, 202, 9
The Godfather Part II, 1974, 96, 9
12 Angry Men, 1957, 195, 9
Schindler's List, 1993, 201, 9
Pulp Fiction, 1994, 154, 8.9
The Lord of the Rings: The Fellowship of the Ring, 2001, 178, 8.8
The Good the Bad and the Ugly, 1966, 178, 8.8
Forrest Gump, 1994, 142, 8.8
Fight Club, 1999, 139, 8.8
The Lord of the Rings: The Two Towers, 2002, 179, 8.8
Inception, 2010, 148, 8.8
Star Wars: Episode V - The Empire Strikes Back, 1980, 124, 8.7
The Matrix, 1999, 136, 8.7
Goodfellas,1990,145,8.7
One Flew Over the Cuckoo's Nest,1975,143,8.7
Se7en,1995,127,8.6
It's a Wonderful Life,1946,130,8.6
```

Figure 4: Data in TXT Format

Copy the text file, and paste in into an Excel spreadsheet. The initial result will not be ideal, since each line in the text file will populate a single cell. We must call up the text import wizard by clicking **Paste Options**, and then **Use Text Import Wizard**.

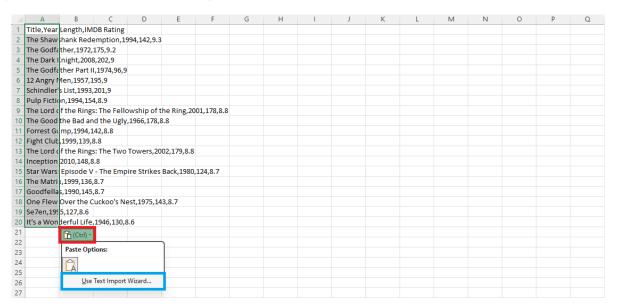


Figure 5: Text File Pasted to Excel

The text import wizard pop-up is shown in Figure 6. Note that in the blue box that the source data is set to Delimited, since the entries are separated by commas. You should choose this same format when each field is separated by tabs as well.<sup>2</sup> This will be the default for most cases when you download a text file with data. Click Next to move along.

	Α	В	С	D	E	F	G	Н	1	J	K	L	М	N	0	Р	Q
1	Title,Year	.ength,IN	1DB Rating														
2	The Shaws	hank Red	emption,1	994,142,9.3	Text Imp	ort Wizar	rd - Step 1 of	3					?	×			
3	The Godfat	her,1972	,175,9.2		The Text V	Vizard has	determined the	at your data is l	Delimited.								
4	The Dark K	night,200	8,202,9					oose the data ty		lossribes vou	er data						
5	The Godfat	her Part	II,1974,96,9				ise next, of the	ouse the data ty	pe mai best	Jescribes you	ir uata.						
6	12 Angry M	en,1957,	195,9			data type											
7	Schindler's	List,1993	3,201,9					scribes your dat									
8	Pulp Fictio	n,1994,15	64,8.9		0	Delimited	<ul> <li>Character</li> </ul>	rs such as comr	nas or tabs se	parate each f	field.						
9	The Lord of	ithe Ring	s: The Fell	owship of the		Fixed widt	h - Fields are	e aligned in colu	imns with spa	ices between	each field.						
10	The Good t	he Bad a	nd the Ugly	,1966,178,8.8													
11	Forrest Gu	mp,1994,	142,8.8		Start imp	ort at row:	1	÷ File orig	in: Windo	ws (ANSI)				~			
12	Fight Club,	1999,139	,8.8														
13	The Lord of	the Ring	s: The Two	Towers, 2002,	,1												
14	Inception,2	2010,148,	8.8		My da	ita has hea	ders.										
15	Star Wars:	Episode \	/ - The Emp	oire Strikes Ba	с												
16	The Matrix	,1999,136	5,8.7		Preview	of selecter	d data:										
17	Goodfellas	,1990,14	5,8.7		1	le Verr	Length, IMD	D. Dating						- I			
18	One Flew (	Over the	Cuckoo's N	est,1975,143,8	3. 2 The	Shawsha	ink Redempt	ion,1994,1	42,9.3					de d			
19	Se7en,1995	6,127,8.6					her, 1972, 17							17.1			
20	It's a Wond	erful Life	,1946,130,	8.6	5 The	Godfath	er Part II	1,1974,96,9									
21							n,1957,195 List,1993										
22																	
23													_				
24									Cancel	<	Back	Next >	Eini	ish			
25					<u> </u>												
26																	

Figure 6: Text Import Wizard Step #1

In this next stage, we can tell Excel that the fields are separated by commas. So, in the red box of Figure 7, deselect Space, and select Comma. Observe how the preview in the blue box changes depending on the selected delimiters.

1	Α	В	С	D	E	E I		G	н	1	J	K	L	М	N	0	Р	Q
1	Title,Year	Length,IN	1DB Rating															
2	The Shaws	hank Red	emption,19	994,142,9.3	3	Text Import	Vizard - S	tep 2 of	3					?	×			
3	The Godfat	ther,1972	,175,9.2		1	This screen lets	vou set th	e delimite	ers vour data o	ontains. You	can see how v	our text is a	iffected in the p	preview below.				
4	The Dark K	night,200	8,202,9		1.0		/		,		,							
5	The Godfa	ther Part	1,1974,96,9			Delimiters												
6	12 Angry N	len,1957,	195,9			<u> </u>		_										
7	Schindler's	5 List,1993	,201,9			Semicolo	n	✓ Treat	t consecutive (	delimiters as o	ne							
	Pulp Fictio					Comma		Text gua	lifier *		~							
			s: The Fello			Space		rext goo										
			nd the Ugly	,1966,178,	8.8	Other:												
	Forrest Gu																	
	Fight Club,																	
			s: The Two	Towers,20	002,1													
	Inception,																	
			/ - The Emp	ire Strikes	Bac	Data preview												
	The Matrix															_		
	Goodfellas					Title, Yea	r, Lengtl	h,IMDB	Rating									
			Cuckoo's Ne	est,1975,14	43,8.	The			Shawshank	,1972,175,		ption,19	94,142,9.3					
	Se7en,199					The			Dark		Knight	t,2008,2	02,9			_		
	It's a Wond	iertul Life	,1946,130,8	5.6		The 12			Godfather Angry		Part Men.1	957,195,	9	II,1974,96	, 9			
21					-	Schindler	's		List,1993	,201,9								
22						4												
23														_				
24										Cancel	<	<u>B</u> ack	<u>N</u> ext >	Ein	ish			
25 26					-			-										

Figure 7: Text Import Wizard Step #2

 $<sup>^{2}</sup>$ The tab key.

	Α	В	С	D	E	F	G	Н	1		J	К	L	М	N	0	Р	Q
1	Title,Year	Length, IN	1DB Rating															
2	The Shaws	hank Red	emption,19	94,142,9.3	Tex	t Import Wiza	ard - Step 3 of	F3						?	×			
3	The Godfat	ther,1972	,175,9.2		This	screen lets vou	select each col	umn and s	et the Data	Format.								
4	The Dark K	night,200	8,202,9			lumn data form												
5	The Godfat	ther Part	II,1974,96,9				ia.											
6	12 Angry M	len,1957,	195,9			General		'Gen	eral' conver	ts numeri	c values to	numbers, dat	e values to da	tes, and all ren	naining			
7	Schindler's	List,1993	3,201,9		C	) Iext		value	es to text.						J.			
8	Pulp Fiction	n,1994,15	4,8.9		C	Date: MDY	~					Advanced						
9	The Lord of	f the Ring	s: The Fello	wship of t	hel 🔿	) Do not impor	t column (skip)											
10	The Good t	he Bad a	nd the Ugly,	,1966,178,	3.8													
11	Forrest Gu	mp,1994,	142,8.8															
12	Fight Club,	1999,139	,8.8															
13	The Lord of	f the Ring	s: The Two	Towers,20	02,1													
14	Inception,2	2010,148,	8.8															
15	Star Wars:	Episode ۱	/ - The Emp	ire Strikes	Bac Da	ta <u>p</u> review												
16	The Matrix	,1999,136	i,8.7												_ 11			
17	Goodfellas	,1990,145	5,8.7			eneral tle			Length	General IMDB Ra	ting							
18	One Flew 0	Over the	Cuckoo's Ne	est,1975,14	3,8. <u>n</u>	ne Shawshan	k Redemptio	on 1994	142	9.3	cing							
	Se7en,1995					he Godfathe he Dark Kni		1972 2008		9.2								
20	It's a Wond	lerful Life	,1946,130,8	.6	Th	e Godfathe	r Part II	1974	96	Ð								
21						Angry Men chindler's		1957 1993		9 9								
22						-									-			
23														_	_			
24									Ca	incel	<	Back	Next >	Ein	ish			
25						_	_	_	_	_				-				
26																		

Figure 8: Text Import Wizard Step #3

In the third and last step of the text import wizard, users can set the data format for each column using the options in the green box. Then, check the preview in blue box, and select Finish. Figure 9 is the resulting table that is generated using the text import wizard.

	А	В	С	D	E	F	G	Н	I	J	K	L
1	Title	Year	Length	IMDB Rati	ng							
2	The Shaws	1994	142	9.3								
3	The Godfa	1972	175	9.2								
4	The Dark H	2008	202	9								
5	The Godfa	1974	96	9								
6	12 Angry I	1957	195	9								
7	Schindler'	1993	201	9								
8	Pulp Fictio	1994	154	8.9								
9	The Lord o	2001	178	8.8								
10	The Good	1966	178	8.8								
11	Forrest Gu	. 1994	142	8.8								
12	Fight Club	1999	139	8.8								
13	The Lord o	2002	179	8.8								
14	Inception	2010	148	8.8								
15	Star Wars:	1980	124	8.7								
16	The Matri	1999	136	8.7								
17	Goodfella	1990	145	8.7								
18	One Flew	1975	143	8.7								
19	Se7en	1995	127	8.6								
20	lt's a Won	1946	130	8.6								
21					<b>E</b>							

Figure 9: Table Generated Using Text Import Wizard

# Topic 2. Power Query: From Web

Instead of manually importing data, we can use Power Query to import, transform, and clean data. We will first examine how to directly import data from websites. Let us return to the Wikipedia article on secondary education attainment. Navigate to the Data tab, and select From Web. Type in the URL in the blue box, and then click OK.

Get Data ~	From Text/CSV From Web From Table/Range Get & Trans	From Pictu  Recent Sou  Existing Co  form Data	ire ~ irces	Refresh	oook Links	ns	Â	Currencies Data Type:	Geography s	~	A Z Z Z Sort	Filter Sort & Filt	Clear Reapply Advanced er
A1 A 1 2 3 4 5 6	From V	C D	E	F G	Н	I	J	K	×	M	N	0	P
7 8 9 10 11 12 13 14	URL 2n.wikiped	ia.org/wiki/List_of	_countries_by_se	econdary_educatio	n_attainment		0	K C	Jancel				

Figure 10: Importing from the Web

Once Excel establishes a link to the webpage, it will open up the Navigator page as shown in Figure 11. To the left hand side, you will find the objects included in the webpage. For this purpose, we should select TableO, and check out a preview of the table in the blue box. If the preview in the blue box indeed matches the table you wish to import, select Transform Data.

A	В	С	D	E	F	G	н	1	J		C L		M	0	P	Q	R	S	T	U	V	W	х	Y	Z	AA	AB	AC
																			-									
					1														×									
					Na	avigator																						
						5																						
										2	Table Vie	/ Web	View															
						Select multip																						
											Table (							G										
					Disp	olay Options	÷			ð	Country		Year	3 to 5 Year	above grad	ation age (%)	1] Ye	ear2										
					- 41	https://e	n.wikipedia	.org/wiki/Li	st_of_cou	ntries	Country		Year	3 to 5 Year	above gradu	tion age (%)	Ye	ear										
						Docum		-			South Kor		2014	99				^										
						Table 0					Georgia		2013	96				- 11	-									
					- L	UII Table 0					Japan		2016	95				- 11										
					_						Croatia		2013	95			21	015	-									
											Ukraine		2012	95														
											Sweden		2013	92			21	015										
											Ireland			null			null 21											
											Armenia		2010	93			11011 21	015										
													2013	94				015										
					1						United Kir		2013	93			-	015										
					1						Kazakhsta	•																
											Poland		2013	83			20	015										
											United Sta	tes	2010	92				_										
											Canada		2010	86				_										
											Greece		2013	92				015	_									
											Slovakia		2013	93				015										
											Cyprus		2013	93				015	-									
											Slovenia		2013	89			20	015										
											Israel		2012	88														
											Russian Fe	deration	2013	87														
											Czech Rep	ublic	2013	90			20	015 💙										
											<						3	>										
															Load -	Transform	Data	Cancel										

Figure 11: Selecting the Data

A new window named Power Query Editor will pop up, which allows the users to edit the data before we import it to Excel. The most basic operations here will be operating on rows and columns.

-ile	r <del>⊊</del> Tab Home		Power Query I nstorm Ad	Editor Id Column View													~
ise & ad ▼	Pefrech	Ad		Choose Remove Columns • Columns	Keep Remove Rows • Rows • Reduce Rows	Ž↓ Ž↓ C Sort	Split Group olumn • By	Data Type: T Use Firs	t Row as H	Headers 🔻	Merge	d Queries 🔻	Manage Parameters • Parameters	Data source settings Data Source	Enter Da	iources <del>-</del> ta	
ries (1)				5													
Table 0				Table.TransformColumnTy											Query Set	tings	
			A <sup>B</sup> C Country	▼ A <sup>II</sup> C Year	✓ <sup>N</sup> C 3 to 5 Years at	ove graduatio	n age (%)[1] 🛛 🔻	A <sup>®</sup> ⊂ Year2	¥ 1	A <sup>II</sup> C 20-24 (%)[2	2] × A	C Year3	✓ A <sup>II</sup> C 20-2	19 (%) [3]	A PROPERTIES		
		1	Country	Year	3 to 5 Years above			Year		20-24	Y	ear	20-29		Name		
					graduation age					(%)			(96)		Table 0		
					(%)												
		2	South Korea	2014	99				null		null 2	015	98		All Properties		
		3	Georgia	2013	96				null		null 2	013	95		▲ APPLIED STEE	PS	
		4	Japan	2016	95				null		null		null				
		5	Croatia	2013	95			2015		95.7	2	013	95		Source		
		6	Ukraine	2012	95				null		null 2	012	94		Navigation × Changed		
		7	Sweden	2013	92			2015		87.3	2	012	94		~ Changed	type	
		8	Ireland		null		nul	2015		92.7			null				
		9	Armenia	2010	93				null		null 2	010	92				
		10	United Kingdom	2013	94			2015		85.7	2	013	92				
		11	Kazakhstan	2010	93				null		null 2	010	92				
		12	Poland	2013	83			2015		90.8	2	013	92				
		13	United States	2010	92				null		null 2	013	91				
		14	Canada	2010	86				null		null 2	010	91				
		15	Greece	2013	92			2015		89.6	2	013	91				
		16	Slovakia	2013	93			2015		91.3	2	013	90				
		17	Cyprus	2013	93			2015		94.3	2	013	90				
			Slovenia	2013	89			2015		90.9	2	013	90				
		19	Israel	2012	88				null		null 2	012	89				
		20	Russian Federation	2013	87				null		null 2		88				
			Czech Republic	2013	90			2015		90.4		013	88				
			Lithuania	2013	91			2015		90.9		013	88				
			France	2013	83			2015		87.2		013	88		-		
		24									-						

Figure 12: Power Query Editor

#### **Editing Columns**

First, to select the columns that are relevant, we can click **Choose Columns** button shown in the red box in Figure 13. Then, you can choose the columns that you would like to have included in the table that will be imported into Excel. You can "uncheck" the items in the blue box that you would not like to have imported.

lle.	▼ =   Table Home	0 - Pov Transf	ver Query Edit orm Add C	or olumn View											/
se & ad ▼	Refrech	Prope	nced Editor ge ▼ Co	Choose Remove	Keep Remove Rows * Rows *	Ą↓ ∡↓	Choose Columns	×		Merge Queries 👻 Append Queries 👻	Manage Parameters •	Data sour settings	5	New Source •	
ose		Query		Manage Columns	Reduce Rows	Sort	Choose the columns to keep			Combine	Parameters	Data Sour	ces	New Query	
es (1)	(	~	√ fx = Tab		s(Data0,{{"Country", t			₿↓		(******					
Table 0		$\times$					(Select All Columns)		ion age	e (%)[1]", type text}, {			×	Query Settings	
		. Mc	Country	✓ A <sup>B</sup> C Year	✓ <sup>№</sup> C 3 to 5 Years at	ove gradu	Country		%)[2]	▼ <sup>A®</sup> c Year3	✓ A <sup>II</sup> C 20-25	9 (%) [3]	*	▲ PROPERTIES	
		1 Co	untry	Year	3 to 5 Years above		✓ Year			Year	20-29			Name	
					graduation age		✓ 3 to 5 Years above graduation age (	6)[1]			(96)			Table 0	
					(%)		✓ Year2								
		2 Sol	uth Korea	2014	99		20-24 (%)[2]			null 2015	98			All Properties	
		3 Ge	orgia	2013	96		✓ Year3			null 2013	95			▲ APPLIED STEPS	
		4 Jap	an	2016	95		20-29 (%) [3]			null	null			Source	
		5 Cro	atia	2013	95		✓ Year4			2013	95			Navigation	
		6 Uk	raine	2012	95		25-29 (%)			null 2012	94			× Changed Type	
		7 Sw	eden	2013	92		Vear5			2012	94			in changed type	
		8 Ire	land		null		25-34 (%)[4]				null				
		9 Arr	menia	2010	93					null 2010	92				
		10 Un	ited Kingdom	2013	94					2013	92				
		11 Ka:	takhstan	2010	93					null 2010	92				
		12 Pol	and	2013	83					2013	92				
		13 Un	ited States	2010	92		1			null 2013	91				
		14 Ca	nada	2010	86					null 2010	91				
		15 Gr	rece	2013	92					2013	91				
		16 Slo	vakia	2013	93					2013	90				
		17 Cyr	orus	2013	93					2013	90				
		18 Slo	venia	2013	89					2013	90				
		19 Isra	oel	2012	88		OK	Cancel		null 2012	89				
		20 Ru	ssian Federation	2013	87					null 2013	88				
		21 Cz	tch Republic	2013	90		2015	90.4		2013	88				
		22 Liti	nuania	2013	91		2015	90.9		2013	88				
		23 Fra	nce	2013	83		2015	87.2		2013	88		~		
		24	<									>			

Figure 13: Power Query Editor: Choosing Columns

#### **Editing Rows**

See the orange box in Figure 14. We removed the Year columns in the previous step, and we can see that the Power Query editor records this change. Choosing the gear icon to the right of each item, you can see the specific changes you made to the imported data. This is a massive improvement over manually editing data.

Next, we can remove rows that are irrelevant for our purposes. For this table, we can see that the variable names are repeated in the first row of the table. We can remove this row by clicking Remove Rows, then selecting Remove Top Rows in the red box. Remove the first row of this table by typing in 1 in the blue box, and click OK.

File			Power Query Editor	mn View									— 🗆	×
oad • Pi	Refresh Preview •	Ad	vanced Editor Cho nage • Colur	nns • Columns • Rows •	Rem	ove AJ Sp	olit Group mn▼ By	Data Type: Text • Use First Ro 1 <sub>92</sub> Replace Valu	w as Headers 🔻	Merge Queries  Append Queries Combine Files	Manage Parameters •	Data source settings	New Source •	
Close		Quer	у ма	nage Columns Reduc				Transform		Combine	Parameters	Data Sources	New Query	
veries (1)	<	$\geq$	√ fx = Table.5	electColumns(#"Changed Type",		Remove Bottom	Rows	tion age (%)[1]", "	20-24 (%)[2]", "20	0-29 (%) [3]", "25-29 (%)",	"25-34 (%)[4]"}	) ~	Query Settings	×
👖 Table 0			ABC Country	A <sup>8</sup> C 3 to 5 Years above graduation a	-	Remove Alternat	e Rows	<sup>A8</sup> c 20-29 (%) [3]	✓ A <sup>8</sup> ⊂ 25-29 (%)	✓ <sup>№</sup> <sub>C</sub> 25-34 (%)[4]	w			
		1	Country	3 to 5 Years above		Remove Duplica	tes	20-29	25-29	25-34		•	<ul> <li>PROPERTIES</li> <li>Name</li> </ul>	
				graduation age				(56)	(96)	(%)			Table 0	
				(%)	Lx	Remove Blank R	ows						All Properties	
		_	South Korea Georgia	99	R.	Remove Errors		98		null	null			
			Georgia Japan	95	120		nuil	95	null	null	null		A APPLIED STEPS	
			Croatia	95	-		nuir		nan	nun			Source	4
		6	Ukraine	95							× null		Navigation	4
		7	Sweden	92		Remove Top Ro	WS						Changed Type	
			Ireland			Specify how many row	s to remove fro	m the top.					➤ Removed Other Columns	Ŷ
			Armenia	93							null			
			United Kingdom	94		Number of rows								
			Kazakhstan	93							null			
		12	Poland	83	1									
		13	United States	92						OK Cancel				
		14	Canada	86	1									
		15	Greece	92	L									
		16	Slovakia	93		91.3		90		null 93				
		17	Cyprus	93		94.3		90		null	null			
		18	Slovenia	89		90.9		90		null 94				
		19	Israel	88			null	89		null 91				
		20	Russian Federation	87			null	88		null 95				
		21	Czech Republic	90		90.4		88		null 94				
		22	Lithuania	91		90.9		88		null 90				
		23	France	83		87.2		88		null 87		~		
		24	Belarus	83			null	88		null	null	~		

Figure 14: Power Query Editor

The other options included in either Keep Rows or Remove Rows may prove quite useful, and we encourage that readers try out these options:

- Keep / Remove Top Rows: Keep / Remove only the top N rows from
- Keep / Remove Bottom Rows: Keep / Remove only the bottom N rows from this table.
- Keep / Remove Range of Rows: Specify the number of rows to keep / remove starting at a specific row.
- Keep / Remove Duplicates: Keep / Remove rows containing duplicated values in the currently selected columns.
- Keep / Remove Errors: Keep / Remove only rows containing errors in the currently selected columns.

#### Loading Data to Excel

Once the table is edited to satisfaction, we can load it to Excel by clicking Close & Load in the red box. It is recommended that users check the Applied Steps in the blue box before loading the table to Excel.

🙂 - ∓   Table   File Home	0 - Power Query Editor Iransform Add Colu	mn View						— 🗆	×
se & Refresh pad ▼ Preview ▼		nns * Columns * Rows * Rows *	Ž↓ Ž↓ Split Group Column ▼ By	Solution 2 Replace Values	Merge Queries  Append Queries Combine Files	Manage Parameters •	Data source settings	New Source   Recent Sources  Enter Data	
lose C	Query Ma	nage Columns Reduce Rows	Sort	Transform	Combine	Parameters	Data Sources	New Query	
eries (1)	$\times \sqrt{f_x}$ = Table.S	<pre>skip(#"Removed Other Columns",1)</pre>					~	Query Settings	×
Table 0	TT, A <sup>B</sup> c Country	ABC 3 to 5 Years above graduation age (%)[1]	A <sup>8</sup> c 20-24 (%)[2]	A <sup>8</sup> ⊂ 20-29 (%) [3] ▼ A <sup>8</sup> ⊂ 25-29 (%)	▼ <sup>A<sup>0</sup>C 25-34 (%)[4]</sup>	w		Query settings	
	1 South Korea	99	null						
		99	กมส		null	null	~	Name	
	2 Georgia	95			null	null		Table 0	
	3 Japan		null 95.7	null 95	null	null		All Properties	
	4 Croatia	95	95.7 null		null	null			
	5 Ukraine		nuii 87.3	94	null null 82	null		APPLIED STEPS	
	6 Sweden	92		94 pull				Source	4
	7 Ireland	93	1 92.7 null		null 91			Navigation	4
	8 Armenia 9 United Kingdom	94		92	null 85	null		Changed Type	
		93	65.7 null					Removed Other Columns	4
	10 Kazakhstan	83		92	null	null		➤ Removed Top Rows	4
	11 Poland				null 94				
	12 United States	92	nuil		null 90				
	13 Canada	86	null		null 93				
	14 Greece	92		91	null 84				
	15 Slovakia	93		90	null 93				
	16 Cyprus	93	94.3	90	null	null			
	17 Slovenia	89		90	null 94				
	18 Israel	88	null		null 91				
	19 Russian Federation	87	null		null 95				
	20 Czech Republic	90	90.4	88	null 94				
	21 Lithuania	91	90.9	88	null 90				
	22 France	83	87.2	88	null 87				
	23 Belarus	83	null		null	null			
	24 Austria	84	88.7	88	null 90				
	25 Finland	85	86.8	87	null 90		~		
	26 Australia	85	null	95	null 88				

Figure 15: Loading Data to Excel

Figure 16 shows the data imported to Excel. The data will automatically be organized as a table as shown in the red box, and the default name will follow the object name we found in Figure 11.

File Home Insert P	age Layout Formulas Data	Review View Automate	Help Table Design	Query								Comments	년 Share
Table_0	ve Duplicates Insert Exp	Properties Properties Properties Properties Properties Properties Properties Properties Properties	Banded Rows	Last Column			r Styles			4 4 3			,
A1 v I X v J	fx Country												
A		B	D	E F	G	H I	 К	L	M N	0	P A		
1 Country	3 to 5 Years above gra	iduation age (%)[1] 💌 20–24 (%	)[2] <b>▼</b> 20–29 (%) [3]									Queries & Connections	V X
2 South Korea	99		98										
3 Georgia	96		95									Queries Connections	
Japan	95												
5 Croatia	95	95.7	95									1 query	
5 Ukraine	95		94									Table 0	
7 Sweden	92	87.3	94	82								121 rows loaded.	
3 Ireland		92.7		91								1211043108060.	
9 Armenia	93		92										
0 United Kingdom	94	85.7	92	85									
1 Kazakhstan	93		92										
2 Poland	83	90.8	92	94									
3 United States	92		91	90									
4 Canada	86		91	93									
5 Greece	92	89.6	91	84									
6 Slovakia	93	91.3	90	93									
17 Cyprus	93	94.3	90										
18 Slovenia	89	90.9	90	94									
19 Israel	88		89	91									
0 Russian Federation	87		88	95									

Figure 16: Imported to Excel

# Topic 3. Power Query: TXT, Splitting, Duplicating, & Grouping

We can also import text files via Power Query. Let us return to the text file we used previously, BUSI201-LEC20-txt. We may import a text file into Excel using the same Power Query framework by selecting From Text/CSV under the Data tab.

File Home Insert Page Layout Formulas	Data Review View Automate	Help
Get From Text/CSV ☐ From Picture ~ Get From Web Data ~ ☐ From Table/Range C Existing Connections	Refresh All • Queries & Connections Properties Workbook Links	Image: Stocks     Currencies     Geography     マ
Get & Transform Data	Queries & Connections	Data Types

Figure 17: Loading Text / CSV to Excel

The window in Figure 18 should pop up when the text file is correctly selected. Since our text file is separated using commas, the delimiter is correctly set to Commas, and the preview in the blue box shows the correct layout for our table. Select Transform Data.

ile Origin	Delimiter				ata Type Detecti	
1252: Western	European (Windows) * Comma			▼ B	lased on first 20	T
Code	Title	Year	Length	Rated	Rating	
IMDB-RANK-01	The Shawshank Redemption	1994	142	R	9.3	
MDB-RANK-02	The Godfather	1972	175	R	9.2	
MDB-RANK-03	The Dark Knight	2008	152	PG-13	9	
MDB-RANK-04	The Godfather Part II	1974	202	R	9	
IMDB-RANK-05	12 Angry Men	1957	96	N/A	9	
MDB-RANK-06	Schindler's List	1993	195	R	9	
MDB-RANK-07	The Lord of the Rings: The Return of the King	2003	201	PG-13	9	
MDB-RANK-08	Pulp Fiction	1994	154	R	8.9	
IMDB-RANK-09	The Lord of the Rings: The Fellowship of the Ring	2001	178	PG-13	8.8	
IMDB-RANK-10	The Good the Bad and the Ugly	1966	178	N/A	8.8	
IMDB-RANK-11	Forrest Gump	1994	142	PG-13	8.8	
IMDB-RANK-12	Fight Club	1999	139	R	8.8	
IMDB-RANK-13	The Lord of the Rings: The Two Towers	2002	179	PG-13	8.8	
IMDB-RANK-14	Inception	2010	148	PG-13	8.8	
IMDB-RANK-15	Star Wars: Episode V - The Empire Strikes Back	1980	124	PG	8.7	
IMDB-RANK-16	The Matrix	1999	136	R	8.7	
IMDB-RANK-17	Goodfellas	1990	145	R	8.7	
IMDB-RANK-18	One Flew Over the Cuckoo's Nest	1975	143	R	8.7	
MDB-RANK-19	Se7en	1995	127	R	8.6	
IMDB-RANK-20	It's a Wonderful Life	1946	130	PG	8.6	

Figure 18: Loading Text / CSV to Excel

Suppose that you want to create a column that splits the first column into many columns that has information on the ranking, which platform the rankings are based on, and the year when the rankings were taken. This can be acheived by splitting the Code column into many parts.

#### **Splitting Columns**

Select Split Column, and then choose By Delimiter, since the Code data is linked via short dashes. There are many different methods to split columns, and those methods may be useful depending on the type of data.

BUSI2	201-LEC20-txt - Power Q Iransform Add Colu	,						—	×
ose & Refresh	Advanced Editor Manage - Cho Colum		Split Split	Row as Headers 🝷	Merge Queries  Append Queries Combine Files	Manage Parameters •	Data source settings	New Source •	
Close (	Query Ma	anage Columns Reduce Rows Sort	By Delimiter		Combine	Parameters	Data Sources	New Query	
eries (1)	× √ fx = Table.T	TransformColumnTypes(#"Promoted Headers",{{"Code",	By Number of Characters	. Int64 Type)	, ("Length", Int64.Type), {	"Rated", type text	). ×	0	
BUSI201-LEC20-txt		- A <sup>2</sup> <sub>C</sub> Title	By Positions	ABc Rated	▼ 1.2 Rating	······		Query Settings	×
			by Positions					▲ PROPERTIES	
	1 IMDB-RANK-01	The Shawshank Redemption	By Lowercase to Uppercas	e 12 R		9.3	~	Name	
	2 IMDB-RANK-02	The Godfather	D. 11	75 R		9.2		BUSI201-LEC20-txt	
	3 IMDB-RANK-03	The Dark Knight	By Uppercase to Lowercas	e i2 PG-13		9		All Properties	
	4 IMDB-RANK-04	The Godfather Part II	By Digit to Non-Digit	12 R 16 N/A		9			
	5 IMDB-RANK-05 6 IMDB-RANK-06	12 Angry Men Schindler's List		IS R		9		APPLIED STEPS	
	7 IMDB-RANK-05	The Lord of the Rings: The Return of the King	By Non-Digit to Digit	JI PG-13		9		Source	4
	8 IMDB-RANK-08	Pulp Fiction	1994	154 R		8.9		Promoted Headers	4
	9 IMDB-RANK-09	The Lord of the Rings: The Fellowship of the Ring	2001	178 PG-13		8.8		➤ Changed Type	
	10 IMDB-RANK-10	The Good the Bad and the Ugly	1966	178 N/A		8.8			
	11 IMDB-RANK-10	Forrest Gump	1994	178 N/A 142 PG-13		8.8			
	12 IMDB-RANK-12	Fight Club	1999	139 R		8.8			
	13 IMDB-RANK-13	The Lord of the Rings: The Two Towers	2002	179 PG-13		8.8			
	14 IMDB-RANK-14	Inception	2010	148 PG-13		8.8			
	15 IMDB-RANK-15	Star Wars: Episode V - The Empire Strikes Back	1980	124 PG		8.7			
	16 IMDB-RANK-16	The Matrix	1999	136 R		8.7			
	17 IMDB-RANK-17	Goodfellas	1990	145 R		8.7			
	18 IMDB-RANK-18	One Flew Over the Cuckoo's Nest	1975	143 R		8.7			
	19 IMDB-RANK-19	Se7en	1995	127 R		8.6			
	20 IMDB-RANK-20	It's a Wonderful Life	1946	130 PG		8.6			
	21 METACRITIC-RANK-01	Citizen Kane	1941	119 PG		100			
	22 METACRITIC-RANK-02	The Godfather	1972	175 R		100			
	23 METACRITIC-RANK-03	Rear Window	1954	112 PG		100			
	24 METACRITIC-RANK-04	Casablanca	1942	102 PG		100			
	25 METACRITIC-RANK-05	Boyhood	2014	165 R		100	~		
	26 METACRITIC-RANK-06	Three Colors: Red	1994	99 R		100	~		

Figure 19: Splitting Columns in Power Query

We can tell Excel which delimiter will be used to split the column in the red box in Figure 20. Set up the options as shown in the red box and blue box to split the column Code. Choosing any of the other options in the blue box will allow the user to split the column in various ways.

			d Column View		^ <b>?</b>
Close &	Refresh	Properties Advanced Editor Manage 🔻	Choose Remove Columns * Columns *	Koon Ramoun A	New Source • Recent Sources • Enter Data
Close	(	Query	Manage Columns	Parameters Data Sources	New Query
Queries [1]	-LEC20-txt		Table.TransformColumnTypes(#	Split Column by Delimiter , ("Rated", type text), v Specify the delimiter used to split the text column.	Query Settings × PROPERTIES Name BUSI201-LEC20-64
		3 IMDB-RANK-03	The Dark Knight	9	
		4 IMDB-RANK-04	The Godfather Part II	9	All Properties
		5 IMDB-RANK-05	12 Angry Men		APPLIED STEPS
		6 IMDB-RANK-06	Schindler's List	O Left-most delimiter 9	Source 🔅
		7 IMDB-RANK-07	The Lord of the Rings: The	Right-most delimiter     g     @ Each occurrence of the delimiter     g	Promoted Headers
		8 IMDB-RANK-08	Pulp Fiction		× Changed Type
		9 IMDB-RANK-09	The Lord of the Rings: The	> Advanced options	
		10 IMDB-RANK-10	The Good the Bad and the	8.8	
		11 IMDB-RANK-11	Forrest Gump	Quote Character 8.8	
		12 IMDB-RANK-12	Fight Club	* * * 88	
		13 IMDB-RANK-13 14 IMDB-RANK-14	The Lord of the Rings: The Inception	8.8 Split using special characters	
		15 IMDB-RANK-19	Star Wars: Episode V - The		
		15 IMDB-RANK-15	The Matrix	Insert special character = 8.7	
		10 IMDB-RANK-10	Goodfellas		
		17 IMDB-RANK-17 18 IMDB-RANK-18	One Flew Over the Cuckor	OK Cancel 8.7	
		19 IMDB-RANK-19	Se7en	8.7	

Figure 20: By Delimiter Options

Clicking OK, the column Code will be split into three parts as shown in the <u>red box</u> in Figure 21. The original column has been split by each occurrence of -, creating columns named Code . 1, Code . 2, and Code . 3. You may double click the header containing the names of the columns to rename the columns.

BUSI2	Irans		wer Query Editor a Column View							— 🗆	×
ose & Refresh pad • Preview •	Mana	nced Editor	Choose Remove Columns • Columns •	Keep Remove Rows • Rows •		Data Type: Text * Split Group 1 Column * By 1 Data Type: Text * Use First Row as Headers *	Merge Queries •	Parameters • se	source ttings	New Source •	
iose (	Query		Manage Columns	Reduce Rows	Sort	Transform	Combine	Parameters Data	Sources	New Query	
eries (1)	$\times$	√ fx =	Table.TransformColumnTyp	es(#"Split Column by [	elimiter"	({"Code.1", type text}, {"Code.2", type text}, {"Code.	3", Int64.Type}})		~	Query Settings	>
BUSI201-LEC20-txt	. 18	Code.1	▼ A <sup>8</sup> C Code-2	<ul> <li>1<sup>2</sup>3 Code.3</li> </ul>	•	A <sup>#</sup> ⊂ Title ▼ 1 <sup>2</sup> 3 Year	✓ 1 <sup>2</sup> <sub>3</sub> Length	✓ A <sup>8</sup> C Rated			
	1 IN	1DB	RANK			The Shawshank Redemption	1994	142 R		A PROPERTIES	
	2 IN	1DB	RANK			The Godfather	1972	175 R	^	BUSI201-LEC20-bit	
	3 IN	IDB	RANK			The Dark Knight	2008	152 PG-13			
	-4 IN	IDB	RANK			The Godfather Part II	1974	202 R		All Properties	
	5 IN	1DB	RANK			12 Angry Men	1957	96 N/A		4 APPLIED STEPS	
	6 IN	1DB	RANK			Schindler's List	1993	195 R			-
	7 IN	1DB	RANK			The Lord of the Rings: The Return of the King	2003	201 PG-13		Source Promoted Headers	
	8 IN	1DB	RANK			Pulp Fiction	1994	154 R		Changed Type	74
	9 IN	1DB	RANK			The Lord of the Rings: The Fellowship of the Ring	2001	178 PG-13		Split Column by Delimiter	4
	10 IN	1DB	RANK		1	The Good the Bad and the Ugly	1966	178 N/A		× Changed Type1	
	11 IN	1DB	RANK		1	Forrest Gump	1994	142 PG-13		V changed type1	
	12 IN	1DB	RANK		1	Fight Club	1999	139 R			
	13 IN	1DB	RANK		1	The Lord of the Rings: The Two Towers	2002	179 PG-13			
	14 IN	1DB	RANK		1	Inception	2010	148 PG-13			
	15 IN	1DB	RANK		1	Star Wars: Episode V - The Empire Strikes Back	1980	124 PG			
	16 IN	IDB	RANK		1	The Matrix	1999	<i>136</i> R			
	17 IN	1DB	RANK		1	Goodfellas	1990	145 R			
	18 IN	1DB	RANK		1	One Flew Over the Cuckoo's Nest	1975	143 R			
	19 IN	IDB	RANK		1	Se7en	1995	127 R			
	20 IN	IDB	RANK		2	It's a Wonderful Life	1946	130 PG			
	21 M	IETACRITIC	RANK			Citizen Kane	1941	119 PG			
	22 M	IETACRITIC	RANK			The Godfather	1972	175 R			
	23 M	IETACRITIC	RANK			Rear Window	1954	112 PG			
	24 M	IETACRITIC	RANK			Casablanca	1942	102 PG			
	25 M	IETACRITIC	RANK			Boyhood	2014	165 R	~		

Figure 21: Splitting Code

# **Creating Duplicate Queries**

We can create duplicate queries by right clicking the original query in the <u>red box</u> in Figure 22, and clicking <u>Duplicate</u>. We will later be using these duplicates to generate new variables, and merge data.

e Hor		-LEC20-txt - Po Instorm Ad	ower Query Ec a Column	litor View								$\sim$
e & Refre d • Previe	C Aresh ew - M	operties Ivanced Editor anage <del>*</del>	Choose R Columns • Co		Keep Remove Rows • Rows •	Ą↓ ∡↓	Column ▼ By 1, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2, 2,	Append Queries	Manage Parameters •	Data source settings	New Source •	
se	Que	ry	Manage Co	lumns	Reduce Rows	Sort	Transform	Combine	Parameters	Data Sources	New Query	
m	< 🔿	< √ fx -	Table.Transform	olumnTypes	(#"Split Column by De	limiter	,{{"Code.1", type text}, {"Code.2", type text}, {"Code.	3", Int64.Type}})		~	Query Settings	
USI201-LEC20-	-txt	, № Code.1	▼ A <sup>®</sup> c Code	2	▼ 1 <sup>2</sup> 3 Code.3		All <sub>C</sub> Title ▼ 1 <sup>2</sup> 3 Year	▼ 1 <sup>2</sup> 3 Length	▼ A <sup>p</sup> c Rate	vit D	Query Settings	
		, WE COULT	- we coue	-2	- 1-3 CODE:3					.u [·	▲ PROPERTIES	
	- Pa	Сору		L			1 The Shawshank Redemption	1994	142 R 175 R	~	Name	
	ĥ	Paste					2 The Godfather	1972 2008	1/5 K 152 PG-13		BUSI201-LEC20-txt	
		Paste		-			The Dark Knight     The Godfather Part II	1974	202 R		All Properties	
	×	Delete					4 The Godfather Part II 5 12 Angry Men	1974	202 R 96 N/A			
				-			6 Schindler's List	1993	195 R		▲ APPLIED STEPS	
	<b>ب</b>	Rename					7 The Lord of the Rings: The Return of the King	2003	201 PG-13		Source	
	Ca.	Duplicate					8 Pulp Fiction	1994	154 R		Promoted Headers	
							9 The Lord of the Rings: The Fellowship of the Ring	2001	178 PG-13		Changed Type	
	C	Reference		-			0 The Good the Bad and the Ugly	1966	178 N/A		Split Column by Delimiter	
		Move To Grou	un t				1 Forrest Gump	1900	142 PG-13		× Changed Type1	
			up ,				2 Fight Club	1999	139 R			
		Move Up		-			3 The Lord of the Rings: The Two Towers	2002	179 PG-13			
		Move Down					4 Inception	2010	148 PG-13			
							5 Star Wars: Episode V - The Empire Strikes Back	1980	124 PG			
		Create Function	on				6 The Matrix	1999	136 R	_		
		Convert To Pa	rameter				7 Goodfellas	1990	145 R			
	_						8 One Flew Over the Cuckoo's Nest	1975	143 R			
	L.	Advanced Ed	itor				9 Se7en	1995	127 R			
		Properties					0 It's a Wonderful Life	1946	130 PG			
		метаските	KANK	1			1 Citizen Kane	1941	119 PG			
	22	METACRITIC	RANK				2 The Godfather	1972	175 R			
	23	METACRITIC	RANK				8 Rear Window	1954	112 PG			
	24	METACRITIC	RANK				4 Casablanca	1942	<i>102</i> PG			
	25	METACRITIC	BANK				5 Boyhood	2014	165 R	~		

Figure 22: Creating Duplicates

#### Grouping

We can use the grouping tool to generate new variables based on this data. Suppose we wanted to know how many movies are in the top movies by its ratings; R, PG, PG-13, etc. Click Group By, and setting up the options as shown in the blue box in Figure 23.<sup>3</sup>

			ower Query Editor ad Column View								× ^ ♥
Load • Pr	lefresh eview •	Advanced Editor	Choose Remove Columns • Columns • I	Keep Remove Rows * Rows *	Split Column	Data Type: Text * Group By J <sub>2</sub> Replace Values	Combine Files	Manage Parameters •	Data source settings	New Source •	
Close		Query	Manage Columns	Reduce Rows	Sort	Transform	Combine	Parameters	Data Sources	New Query	
Queries [1]	<	× √ /x -	Table.TransformColumnTypes(#"	Split Column by Deli	miter".{{"Code.1"	, type text}, {"Code.2", type text}, {"Code	.3". Int64.Type}})		~	Query Settings	×
A BUSI201-LEG	C20-txt	III. NC Code.1	▼ A <sup>B</sup> c Code.2	* 123 Code.3	▼ A <sup>0</sup> c Title	▼ 1 <sup>2</sup> 3 Year	▼ 1 <sup>2</sup> 3 Length	▼ A <sup>ll</sup> c Rate	4	Query settings	^
		1 IMDB	RANK	13 00000	inc mic	13 100	×	142 B		▲ PROPERTIES	
		2 IMDB	RANK	Course Day				175 R	^	Name	
		3 IMDB	RANK	Group By				152 PG-13		BUSI201-LEC20-txt	
		4 IMDB	RANK	Specify the column	n to group by and th	e desired output.		202 R	_	All Properties	
		5 IMDB	RANK	⊕ Basic ○ Advan	ced			96 N/A		A APPLIED STEPS	
		6 IMDB	RANK					195 R			
		7 IMDB	RANK					201 PG-13		Source	*
		8 IMDB	RANK	Rated	Ť			154 R	_	Promoted Headers Changed Type	8
		9 IMDB	RANK	New column name	Op	eration Column		178 PG-13		Split Column by Delimiter	8
		10 IMDB	RANK	Count		ount Rows *	· ·	178 N/A		× Changed Type1	74
		11 IMDB	RANK					142 PG-13		r changed typer	
		12 IMDB	RANK					139 R			
		13 IMDB	RANK				OK Cancel	179 PG-13			
		14 IMDB	RANK					148 PG-13	_		
		15 IMDB	RANK					124 PG			
		16 IMDB	RANK		16 The Matrix		1999	136 R			
		17 IMDB	RANK		17 Goodfellas		1990	145 R			
		18 IMDB	RANK		18 One Flew O	ver the Cuckoo's Nest	1975	143 R			
		19 IMDB	RANK		19 Se7en		1995	127 R			
		20 IMDB	RANK		20 It's a Wonde	erful Life	1946	130 PG			
		21 METACRITIC	RANK		1 Citizen Kane		1941	119 PG			
		22 METACRITIC	RANK		2 The Godfath		1972	175 R			
		23 METACRITIC	RANK		3 Rear Windo	W	1954	112 PG			
		24 METACRITIC	RANK		4 Casablanca		1942	102 PG	~		
		25 METACRITIC	RANK		5 Boyhood		2014	<i>165</i> R	*		
R COLUMNIC 40 DOM		26 <						_	>	2050/00/00/00/00/00/00/00/00/00/00/00/00/	

Figure 23: Grouping by Rated

Figure 24 shows us how the data will be transformed following the grouping described above. We will later see how we can merge query tables to consolidate multiple data sources.

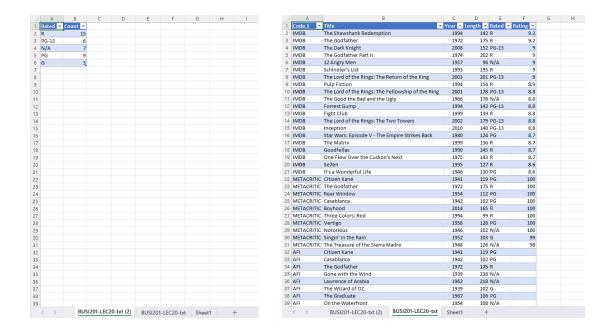
File	• <del>-</del>   Cou Home	nt by Ratings - Pov Iranstorm Ad	ver Query I Id Column	ditor View									- 0	×
Close & Load ▼	Refresh	Properties	Choose Columns •	Remove Columns *	Keep Remov Rows • Rows				Data Type: Whole Number * Use First Row as Headers * <sup>1</sup> / <sub>22</sub> Replace Values	Merge Queries  Append Queries Combine Files	Manage Parameters *	Data source settings	New Source •	
Close		Query	Manage	Columns	Reduce Rows	Sort			Transform	Combine	Parameters	Data Sources	New Query	
Queres (2)				#"Changed Type	1, ("Bated"), ((" 12 15 6 7 9 3 3		ch Table.RowCou	unt(_), i					Query Settings  PROPERTIES Name Court by Ratings  All Properties  APPLID STFDS  Source Promotel Headers Changed Type Self Column by Deminer Changed Type1  X Grouped Rows	×

Figure 24: Grouped by Rated

<sup>&</sup>lt;sup>3</sup>At this point, we ignore the issue of duplicates. For instance, The Godfather is included in all three lists.

# Topic 4. Power Query: Refreshing Data

You might still believe that manually importing data isn't too troublesome, considering the need to learn another tool. However, data imported using Power Query offers a crucial advantage over manual imports – it enables us to refresh the tables in Excel when the source data changes. Import the table using Power Query from **Topic 3** to obtain the two worksheets depicted below.



Suppose now that the rating for the top-rated movie on IMDB, "The Shawshank Redemption," has been updated to 9.5. Additionally, let's assume that "The Dark Knight" has had its PG-13 rating updated to an R rating. Open the file BUSI201-LEC20-txt, update the ratings in the source file, and then save the file.

	BUSI201-LEC20-txt	٠	+
File	Edit View		
IMDE IMDE IMDE IMDE	2,Title,Year,Length,Rated 3-RANK-01,The Shawshank R 3-RANK-02,The Godfather,1 3-RANK-03,The Dark Knight 3-RANK-04,The Godfather P 3-RANK-05,12 Angry Men,19	edemp 972,1 ,2008 art I	tion,1994,142,R <mark>9.5</mark> 75,R,9.2 ,152 <mark>R</mark> 9 I,1974,202,R,9

Figure 25: Updating Source Data

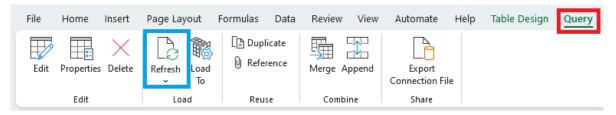


Figure 26: Refreshing Tables

Return to the Excel worksheet BUSI201-LEC20-txt containing the main source, click the tables generated using Power Query to make the Query tab available. Then, click Refresh to update the table.

	А	В	С	D	Е	F	G	Н
1	Code.1	Title	💌 Year 💌	Length 💌	Rated 💌	Rating 💌		
2	IMDB	The Shawshank Redemption	1994	142	R	9.5		
3	IMDB	The Godfather	1972	175	R	9.2		
4	IMDB	The Dark Knight	2008	152	R	9		
5	IMDB	The Godfather Part II	1974	202	R	9		
6	IMDB	12 Angry Men	1957	96	N/A	9		
7	IMDB	Schindler's List	1993	195	R	9		

Figure 27: First Table Refreshed

See Figure 27 to observe the refreshed table. Follow the same workflow to update the second table, where we grouped the data. Once you refresh the second table in the worksheet BUSI201-LEC20-txt(2), you can see that the grouped table displays the updated distribution as shown in Figure 28.

	А	В	С	D	E
1	Rated 💌	Count 💌			
2	R	16			
3	N/A	7			
4	PG-13	5			
5	PG	9			
6	G	3			
7					

Figure 28: Second Table Refreshed

While we cannot demonstrate here, if information on a webpage is updated, you may update the contents of your Excel spreadsheet by clicking refresh.