BDD200NAA - Project 02

Ran Arino

1. Functional dependencies

The following steps show the normalization step, in other words, how I changed one large table into normalized tables.

First Normal Form(1NF) Dependency Diagram

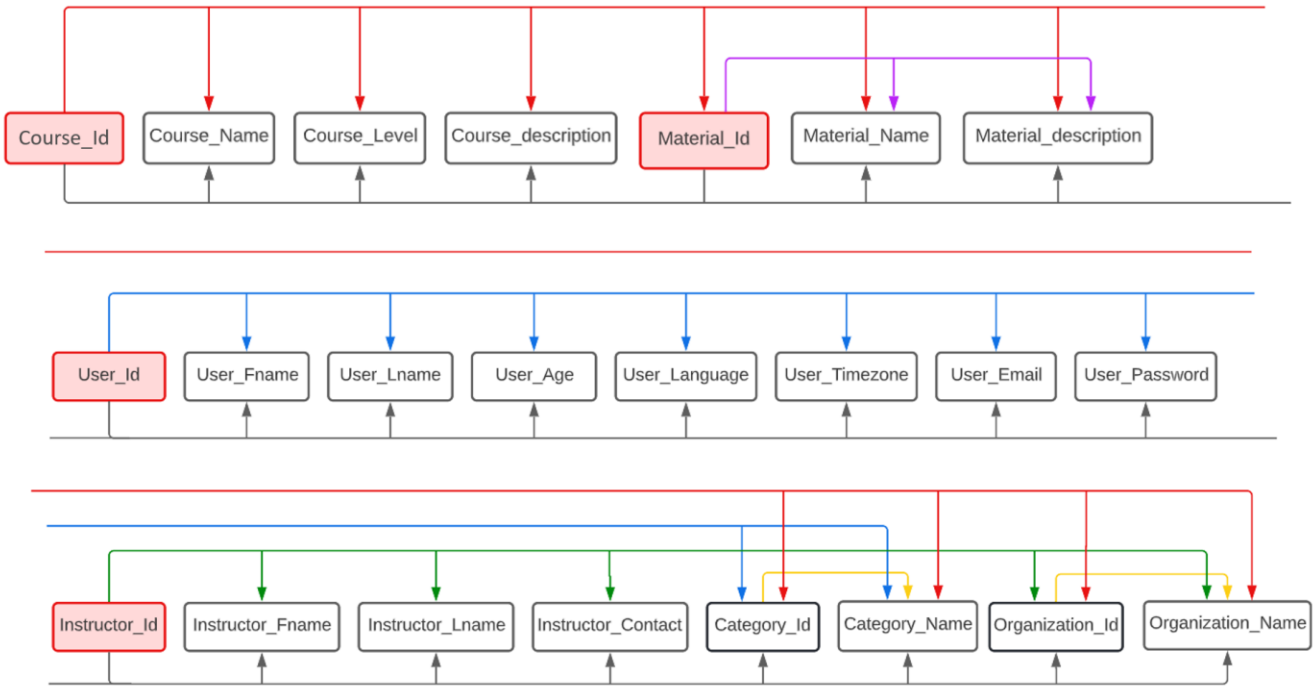


Table Name: **E_LEARNING**

(Course_Id, Material_Id, User_Id, Instructor_Id, Course_Name, Course_Level, Course_Description, Material_Name, Material_Description, User_Fname, User_Lname, User_Age, User_Language, User_Timezone, User_Email, User_Passoword, Instructor_Fname, Instructor_Lname, Instructor_Contract, Category_Id, Category_Name, Organization_Id, Organization_Name)

Partial Dependencies

- (Course_Id --> Course_Name, Course_Level, Course_Description, Category_Id, Category_Name, Organization_Id, Organization_Name)

- (Material_Id --> Material_Name, Material_Description)
- (User_Id --> User_Fname, User_Lname, User_Age, User_Language, User_Timezone, User_Email, User_Passoword, Category_id, Category_name)
- (Instructor_Id --> Instructor_Fname, Instructor_Lname, Instructor_Contract, Organization_Id, Organization_Name)

Transitive Dependencies

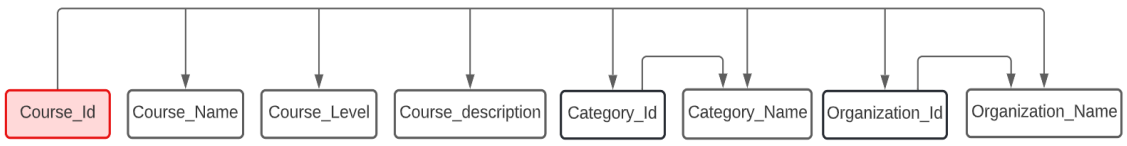
- (Category_Id --> Category_Name)
- (Organization_Id --> Organization_Name)

Second Normal Form(2NF) Dependency Diagram

Table Name: **COURSE**

	A	B	C	D	E	F	G	H
1	COURSE_ID	COURSE_NAME	COURSE_LEVEL	COURSE_DESCRIPTION	CATEGORY_ID	CATEGORY_NAME	ORGANIZATION_ID	ORGANIZATION_NAME
2	1	Google Data Science	Beginner	The course, Google Data Sci	1	Data Science	1	Google
3	2	Data Science on Goog	Advanced	The course, Data Science on	1	Data Science	1	Google
4	3	How to Use Google Cl	Intermediate	The course, How to Use Goo	2	Business	1	Google
5	4	Business Strategies	Advanced	The course, Business Strateg	2	Business	1	Google
6	5	Introduction to Progr	Beginner	The course, Introduction to	3	Computer Science	1	Google
7	6	IBM Data Science	Beginner	The course, IBM Data Scienc	1	Data Science	2	IBM
8	7	Applied Data Science	Intermediate	The course, Applied Data Sc	1	Data Science	2	IBM
9	8	Introduction to Busin	Beginner	The course, Introduction to	2	Business	2	IBM
10	9	Business Analytics	Advanced	The course, Business Analyt	2	Business	2	IBM
11	10	Computer Science	Intermediate	The course, Computer Scien	3	Computer Science	2	IBM
12	11	Microsoft Data Scienc	Beginner	The course, Microsoft Data	1	Data Science	3	Microsoft
13	12	How to use Excel for D	Beginner	The course, How to use Exce	1	Data Science	3	Microsoft
14	13	Business Managemen	Intermediate	The course, Business Manag	2	Business	3	Microsoft
15	14	Business Analytics	Advanced	The course, Business Analyt	2	Business	3	Microsoft
16	15	Introduction to Comp	Beginner	The course, Introduction to	3	Computer Science	3	Microsoft
17	16	Machine Learning Me	Intermediate	The course, Machine Learni	1	Data Science	4	University of Toronto
18	17	How to launch New B	Advanced	The course, How to launch M	2	Business	4	University of Toronto
19	18	Coding: C and C++	Intermediate	The course, Coding: C and C	3	Computer Science	4	University of Toronto
20	19	Introduction to Linea	Beginner	The course, Introduction to	4	Math	4	University of Toronto
21	20	Mathematics for Mac	Intermediate	The course, Mathematics fo	4	Math	4	University of Toronto
22	21	Introduction to Data S	Beginner	The course, Introduction to	1	Data Science	5	University of Michigan

- Cell A is unique and determines Cell B, C, and D.
- “Cell E & F” and “Cell G & H” are repeated, and they have a determinant-dependence relationship. It means that both have a transitive dependency.



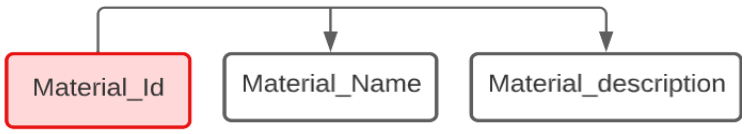
COURSE (Course_Id, Course_Name, Course_Level, Course_Description, Category_Id, Category_Name, Organization_Id, Organization_Name)

Transitive Dependency
- (Category_Id --> Category_Name)
- (Organization_Id --> Organization_Name)

Table Name: **MATERIAL**

	A	B	C
1	MATERIAL	MATERIAL_NAME	MATERIAL_DESCRIPTION
2	1	Google Data Science - Step 1	This is the first part of the Google Data Sc
3	2	Google Data Science - Step 2	This is the second part of the Google Dat
4	3	Google Data Science - Step 3	This is the third part of the Google Data S
5	4	Data Science on Google Cloud - Part 1	This is the first part of the Data Science o
6	5	Data Science on Google Cloud - Part 2	This is the second part of the Data Scienc
7	6	Data Science on Google Cloud - Part 3	This is the third part of the Data Science
8	7	How to Use Google Cloud on Business - Step 1	This is the first part of the How to Use Gc
9	8	How to Use Google Cloud on Business - Step 2	This is the second part of the How to Use
10	9	How to Use Google Cloud on Business - Step 3	This is the third part of the How to Use G
11	10	Business Strategies - First Part	This is the first part of the Business Strate
12	11	Business Strategies - Second Part	This is the second part of the Business St
13	12	Business Strategies - Third Part	This is the third part of the Business Strat
14	13	Introduction to Programming Languages - Class 1	This is the first part of the Introduction t
15	14	Introduction to Programming Languages - Class 2	This is the second part of the Introductio
16	15	Introduction to Programming Languages - Class 3	This is the third part of the Introduction t
17	16	Introduction to Programming Languages - Class 4	This is the forth part of the Introduction
18	17	IBM Data Science - Class 01	This is the first part of the IBM Data Scier
19	18	IBM Data Science - Class 02	This is the second part of the IBM Data Sc
20	19	IBM Data Science - Class 03	This is the third part of the IBM Data Scie

- Cell A (MATERIAL_ID) is obviously unique and determines Cell B and C.
- There are no transitive dependencies. It means that this table has already been converted into 3NF.



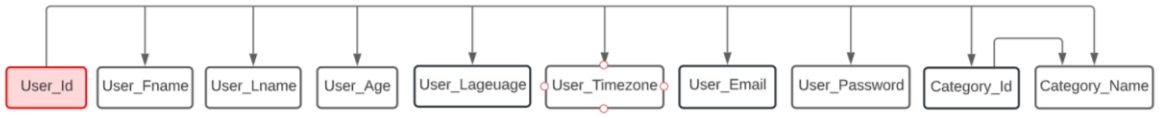
MATERIAL (Material_Id, Material_Name, Material_Desscription)

Table Name: **USERS**

	A	B	C	D	E	F	G	H	I	J
1	USER_ID	USER_FNAME	USER_LNAME	USER_AGE	USER_LANGUAGE	USER_TIMEZONE	USER_EMAIL	USER_PASSWORD	CATEGORY_ID	CATEGORY_NAME
2	1	Kyle	Schneider	27	English	Anchorage (AKDT)	kyle.schneider(	gEWLVSYcfWg2.?9	2	Business
3	2	Kyle	Martin	18	English	Phoenix (MST)	kyle.martin@ei	cWnZcL2+5w/szwS	4	Math
4	3	Meg	Derek	33	English	Honolulu (HST)	meg.derek@fli	,GGPt3+Ktq*TqEE	4	Math
5	4	Shelley	Peckinpah	38	English	Chicago (CDT)	shelley.peckin	uugjxAHZoeh(,Jq	1	Data Science
6	5	Prem	Garcia	32	English	Los Angeles (PDT)	prem.garcia@j	q/+@s1yWGjX2Q3A	2	Business
7	6	Bo	Hitchcock	41	English	Chicago (CDT)	bo.hitchcock@	DgJv?p/p95xVvzn	3	Computer Science
8	7	Bob	McCarthy	25	English	Denver (MDT)	bob.mccarthy(	6EIB+(pUme@Ytb9	5	Health
9	8	Dom	McQueen	28	English	Chicago (CDT)	dom.mcqueen	ciJMX7w5gLfNBzB	3	Computer Science
10	9	Dom	Hoskins	24	English	Honolulu (HST)	dom.hoskins@	QMAH4ekUA*KA)dG	1	Data Science
11	10	Don	Siegel	37	English	Honolulu (HST)	don.siegel@bit	)Z-v3H.)k48?Bxr	1	Data Science
12	11	Scott	Jordan	29	Hindi	Chicago (CDT)	scott.jordan@v	qRNO51Qo@tFqYmB	1	Data Science
13	12	Shammi	Pacino	31	Hindi	Honolulu (HST)	shammi.pacinc	E+CSi7APP3bXyXN	2	Business
14	13	Sharmila	Kazan	32	Hindi	Phoenix (MST)	sharmila.kazan	*uq8h)mSxOWxybB	1	Data Science
15	14	Sharmila	Fonda	34	Hindi	Chicago (CDT)	sharmila.fonda	piN@A6WPVzGYuI5	4	Math
16	15	Shelley	Taylor	31	Hindi	Honolulu (HST)	shelley.taylor@	YLEp6r?)9.UIiS,	1	Data Science
17	16	Shyam	Plummer	29	Hindi	Denver (MDT)	shyam.plumme	S5?ctk4ZlEOk+@*	3	Computer Science
18	17	Silk	Kurosawa	24	Hindi	Honolulu (HST)	silk.kurosawa@	273bpo1@fd7G2Mw	1	Data Science
19	18	Sivaji	Gielgud	21	Hindi	Denver (MDT)	sivaji.gielgud@	n.FwcN1Y-zqpbe6	3	Computer Science
20	19	M. Emmet	Stockwell	34	Italian	Denver (MDT)	m.emmet.stocl	72ZL/2I4XbzMn5d	2	Business
21	20	M. Emmet	Olin	38	Italian	Central European T	m.emmet.olin@	uC52Jixm,vm.KoP	3	Computer Science
22	21	Malcolm	Field	21	Italian	Honolulu (HST)	malcolm.field@	m1OzRYZPP-)J7ud	3	Computer Science
23	22	Mammutti	Sutherland	26	Italian	Chicago (CDT)	mammutti.sutl	i4+6BGiVJkY1QsT	4	Math
24	23	Mani	Kazan	41	Italian	Central European T	mani.kazan@tr	LYwlEDrQ9A-92jn	4	Math
25	24	Mani	Buckley	35	Italian	Los Angeles (PDT)	mani.buckley@	9LUVsc.luTP8tMq*	4	Math

report (2)

- Cell *A* is not duplicated and determines Cell *B*-H.
- Cell *I* determines Cell *J* , so they are transitive dependencies.



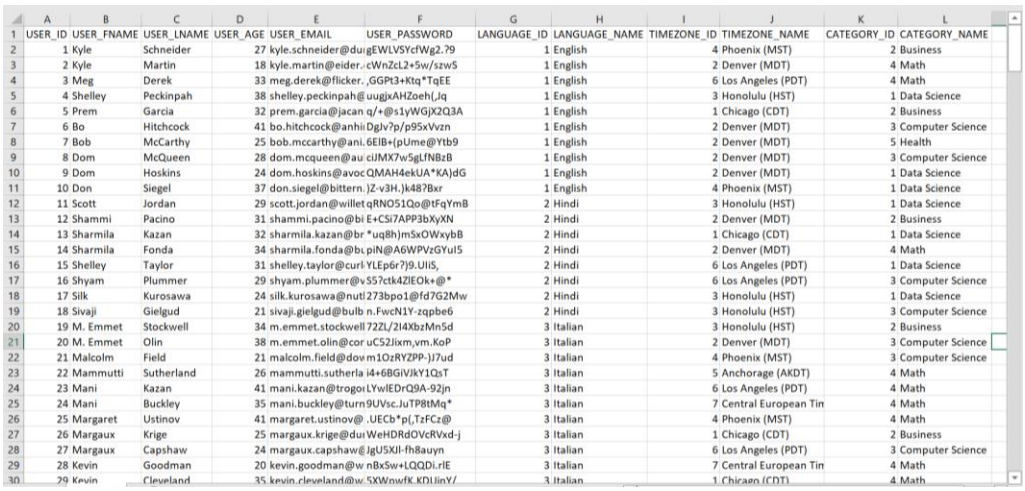
USERS (User_Id, User_Fname, User_Lname, User_Age, User_Language, User_Timezone, User_Email, User_Passoword, Category_Id, Category_Name)

Transitive Dependency
- (Category_Id --> Category_Name)

However, the two columns, USER_LANGUAGE and USER_TIMEZONE, repeat the same variable name many times. In addition, their name is likely to change, so it may cause an update error. More specifically, suppose an organization decides to change the name of a certain time zone globally, such as “Japanese Standard Time (JST)” --> “Tokyo Standard Time (TST)”. In that case, it will consume a considerable amount of time to apply this change to the E-Learning database (USERS table). It is because we need to change its name one by one. Also, the column “COURSE_LEVEL” repeats the same variable name many times, but it’s not likely to happen the update anomalies. It is because the names (like "Advanced") are arbitrary variables. In other words, these names are defined by the database developers, so developers don’t need to change their names broadly, such as “Advanced" --> “Advance”.

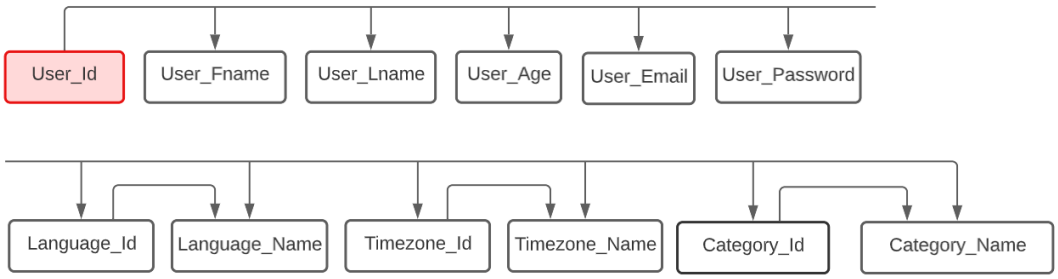
To sum up, in order to avoid the possible update anomalies against the USER_LANGUAGE and USER_TIMEZONE, I decided to rewrite the 2NF and update the table structure like below.

Table Name: **USERS (revised version)**



	A	B	C	D	E	F	G	H	I	J	K	L
1	USER_ID	USER_FNAME	USER_LNAME	USER_AGE	USER_EMAIL	USER_PASSWORD	LANGUAGE_ID	LANGUAGE_NAME	TIMEZONE_ID	TIMEZONE_NAME	CATEGORY_ID	CATEGORY_NAME
2	1	Kyle	Schneider	27	kyle.schneider@du	gEWLVSYcfWg2.?9	1	English	4	Phoenix (MST)	2	Business
3	2	Kyle	Martin	18	kyle.martin@eider.	cWnZcL2+5w/szw5	1	English	2	Denver (MDT)	4	Math
4	3	Meg	Derek	33	meg.derek@flicker.	,GGPt3+Ktq*TqEE	1	English	6	Los Angeles (PDT)	4	Math
5	4	Shelley	Peckinpah	38	shelley.peckinpah@	uugjxAHZoeh(,Jq	1	English	3	Honolulu (HST)	1	Data Science
6	5	Prem	Garcia	32	prem.garcia@jacan	q/+@s1yWGjX2Q3A	1	English	1	Chicago (CDT)	2	Business
7	6	Bo	Hitchcock	41	bo.hitchcock@anhi	DgJv?p/p95xVvzn	1	English	2	Denver (MDT)	3	Computer Science
8	7	Bob	McCarthy	25	bob.mccarthy@ani.	6EIB+(pUme@Ytb9	1	English	2	Denver (MDT)	5	Health
9	8	Dom	McQueen	28	dom.mcqueen@au	ciJMX7w5gLfNBzB	1	English	2	Denver (MDT)	3	Computer Science
10	9	Dom	Hoskins	24	dom.hoskins@avoc	QMAH4ekUA*KA)dG	1	English	2	Denver (MDT)	1	Data Science
11	10	Don	Siegel	37	don.siegel@bittern.	)Z-v3H.)k48?Bxr	1	English	4	Phoenix (MST)	1	Data Science
12	11	Scott	Jordan	29	scott.jordan@willet	qRNO51Qo@tFqYmB	2	Hindi	3	Honolulu (HST)	1	Data Science
13	12	Shammi	Pacino	31	shammi.pacino@bi	E+CSi7APP3bXyXN	2	Hindi	2	Denver (MDT)	2	Business
14	13	Sharmila	Kazan	32	sharmila.kazan@br	*uq8h)mSxOWxybB	2	Hindi	1	Chicago (CDT)	1	Data Science
15	14	Sharmila	Fonda	34	sharmila.fonda@bu	piN@A6WPVzGYuI5	2	Hindi	2	Denver (MDT)	4	Math
16	15	Shelley	Taylor	31	shelley.taylor@curl	YLEp6r?)9.UIiS,	2	Hindi	6	Los Angeles (PDT)	1	Data Science
17	16	Shyam	Plummer	29	shyam.plummer@v	55?ctk4ZlEOk+@*	2	Hindi	6	Los Angeles (PDT)	3	Computer Science
18	17	Silk	Kurosawa	24	silk.kurosawa@nutl	273bpo1@fd7G2Mw	2	Hindi	3	Honolulu (HST)	1	Data Science
19	18	Sivaji	Gielgud	21	sivaji.gielgud@bulb	n.FwcN1Y-zqpbe6	2	Hindi	3	Honolulu (HST)	3	Computer Science
20	19	M. Emmet	Stockwell	34	m.emmet.stockwell	72ZL/2I4XbzMn5d	3	Italian	3	Honolulu (HST)	2	Business
21	20	M. Emmet	Olin	38	m.emmet.olin@cor	uC52Jixm,vm.KoP	3	Italian	2	Denver (MDT)	3	Computer Science
22	21	Malcolm	Field	21	malcolm.field@dov	m1OzRYZPP-)J7ud	3	Italian	4	Phoenix (MST)	3	Computer Science
23	22	Mammutti	Sutherland	26	mammutti.sutherla	i4+6BGiVJkY1QsT	3	Italian	5	Anchorage (AKDT)	4	Math
24	23	Mani	Kazan	41	mani.kazan@trogo	LYwlEDrQ9A-92jn	3	Italian	6	Los Angeles (PDT)	4	Math
25	24	Mani	Buckley	35	mani.buckley@turn	9UVsc.JuTP8tMq*	3	Italian	7	Central European Tin	4	Math
26	25	Margaret	Ustinov	41	margaret.ustinov@	.UECb*p(,TzFCz@	3	Italian	4	Phoenix (MST)	4	Math
27	26	Margaux	Krige	25	margaux.krige@du	WeHDRdOVcRVxd-j	3	Italian	1	Chicago (CDT)	2	Business
28	27	Margaux	Capshaw	24	margaux.capshaw@	JgU5XJI-fh8auyn	3	Italian	6	Los Angeles (PDT)	3	Computer Science
29	28	Kevin	Goodman	20	kevin.goodman@w	nBxSw+LQQDi.rlE	3	Italian	7	Central European Tin	4	Math
30	29	Kevin	Cleveland	35	kevin.cleveland@w	5XWnwfK.KDLiinY/	3	Italian	1	Chicago (CDT)	4	Math

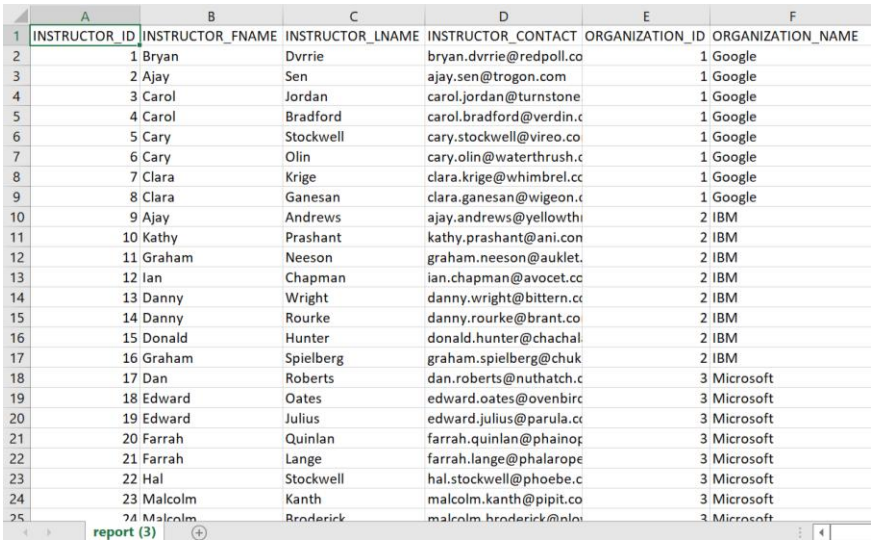
As we can see, Cell *G-H*, Cell *I-J*, and Cell *K-L* are a determine-dependent relationship. For example, LANGUAGE_ID = '1' corresponds with LANGUAGE_NAME = 'English'. It means that they are transitive dependencies.



USERS (<u>User_Id</u>, User_Fname, User_Lname, User_Age, User_Email, User_Passoword, Language_Id, Language_Name, Timezone_Id, Timezone_Name, Category_Id, Category_Name)

Transitive Dependency
- (Language_Id --> Language_Name)
- (Timezone_Id --> Timezone_Name)
- (Category_Id --> Category_Name)

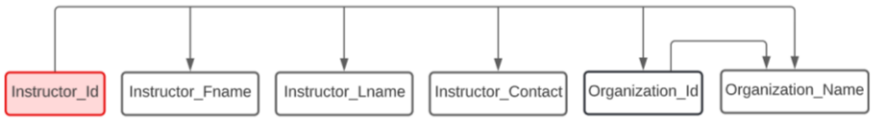
Table Name: **INSTRUCTOR**



	A	B	C	D	E	F
1	INSTRUCTOR_ID	INSTRUCTOR_FNAME	INSTRUCTOR_LNAME	INSTRUCTOR_CONTACT	ORGANIZATION_ID	ORGANIZATION_NAME
2	1	Bryan	Dvrrie	bryan.dvrrie@redpoll.co	1	Google
3	2	Ajay	Sen	ajay.sen@trogon.com	1	Google
4	3	Carol	Jordan	carol.jordan@turnstone	1	Google
5	4	Carol	Bradford	carol.bradford@verdin.c	1	Google
6	5	Cary	Stockwell	cary.stockwell@vireo.co	1	Google
7	6	Cary	Olin	cary.olin@waterthrush.c	1	Google
8	7	Clara	Krige	clara.krige@whimbrel.cc	1	Google
9	8	Clara	Ganesan	clara.ganesan@wigeon.c	1	Google
10	9	Ajay	Andrews	ajay.andrews@yellowthi	2	IBM
11	10	Kathy	Prashant	kathy.prashant@ani.con	2	IBM
12	11	Graham	Neeson	graham.neeson@auklet.	2	IBM
13	12	Ian	Chapman	ian.chapman@avocet.cc	2	IBM
14	13	Danny	Wright	danny.wright@bittern.cc	2	IBM
15	14	Danny	Rourke	danny.rourke@brant.co	2	IBM
16	15	Donald	Hunter	donald.hunter@chachal	2	IBM
17	16	Graham	Spielberg	graham.spielberg@chuk	2	IBM
18	17	Dan	Roberts	dan.roberts@nuthatch.c	3	Microsoft
19	18	Edward	Oates	edward.oates@ovenbirc	3	Microsoft
20	19	Edward	Julius	edward.julius@parula.cc	3	Microsoft
21	20	Farrah	Quinlan	farrah.quinlan@phainop	3	Microsoft
22	21	Farrah	Lange	farrah.lange@phalarope	3	Microsoft
23	22	Hal	Stockwell	hal.stockwell@phoebe.c	3	Microsoft
24	23	Malcolm	Kanth	malcolm.kanth@pipit.co	3	Microsoft
25	24	Malcolm	Broderick	malcolm.broderick@plov	3	Microsoft

report (3)

- Cell A determines Cell B, C, and D. Also, the id is uniquely identifiable.
- Cell E determines Cell F. Also, they are duplicated many times in a table. Thus, these are transitive dependencies.



INSTRUCTOR (Instructor_Id, Instructor_Fname, Instructor_Lname, Instructor_Contract, Organization_Id, Organization_Name)

Transitive Dependency
- (Organization_Id --> Organization_Name)

Third Normal Form(3NF) Dependency Diagram

Table Name: **COURSE**

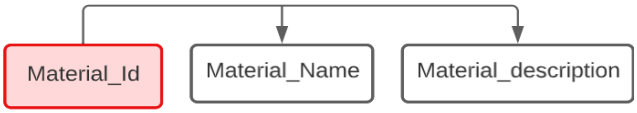
	A	B	C	D	E	F	G
1	COURSE_ID	COURSE_NAME	COURSE_LEVEL	COURSE_DESCRIPTION	ORGANIZATION_ID	CATEGORY_ID	
2	1	Google Data Science	Beginner	The course, Google Data Science, is offered by Goog	1	1	
3	2	Data Science on Google Cloud	Advanced	The course, Data Science on Google Cloud, is offere	1	1	
4	3	How to Use Google Cloud on Business	Intermediate	The course, How to Use Google Cloud on Business,	1	2	
5	4	Business Strategies	Advanced	The course, Business Strategies, is offered by Googl	1	2	
6	5	Introduction to Programming Languages	Beginner	The course, Introduction to Programming Language	1	3	
7	6	IBM Data Science	Beginner	The course, IBM Data Science, is offered by IBM, an	2	1	
8	7	Applied Data Science with Python	Intermediate	The course, Applied Data Science with Python, is off	2	1	
9	8	Introduction to Business	Beginner	The course, Introduction to Business, is offered by I	2	2	
10	9	Business Analytics	Advanced	The course, Business Analytics, is offered by IBM, ar	2	2	
11	10	Computer Science	Intermediate	The course, Computer Science, is offered by IBM, ar	2	3	
12	11	Microsoft Data Science	Beginner	The course, Microsoft Data Science, is offered by M	3	1	
13	12	How to use Excel for Data Analysis	Beginner	The course, How to use Excel for Data Analysis, is of	3	1	
14	13	Business Management	Intermediate	The course, Business Management, is offered by Mi	3	2	
15	14	Business Analytics	Advanced	The course, Business Analytics, is offered by Micros	3	2	
16	15	Introduction to Computer Science and Programming	Beginner	The course, Introduction to Computer Science and F	3	3	
17	16	Machine Learning Methods	Intermediate	The course, Machine Learning Methods, is offered b	4	1	
18	17	How to launch New Business	Advanced	The course, How to launch New Business, is offered	4	2	
19	18	Coding: C and C++	Intermediate	The course, Coding: C and C++, is offered by Univers	4	3	
20	19	Introduction to Linear Algebra	Beginner	The course, Introduction to Linear Algebra, is offere	4	4	
21	20	Mathematics for Machine Learning	Intermediate	The course, Mathematics for Machine Learning, is o	4	4	
22	21	Introduction to Data Science	Beginner	The course, Introduction to Data Science, is offered	5	1	
23	22	Project Management Principles		The course, Project Management Principles, is offer	5	2	
24	23	How to create Algorithm in Python	Intermediate	The course, How to create Algorithm in Python, is of	5	3	
25	24	History of Mathematics	Beginner	The course, History of Mathematics, is offered by U	5	4	
26	25	Linear Regression	Beginner	The course, Linear Regression, is offered by Univers	5	4	
27	26	Introduction to Statistics	Beginner	The course, Introduction to Statistics, is offered by S	6	4	
28	27	Statistical Inference	Intermediate	The course, Statistical Inference, is offered by Stanf	6	4	
29	28	Mathematical Biostatistics	Advanced	The course, Mathematical Biostatistics, is offered by	6	4	
30	29	Discrete Mathematics	Beginner	The course, Discrete Mathematics, is offered by Sta	6	4	

Considering the above table, we can confirm that there is only one determinant (COURSE_ID), and there are no dependencies. Thus, the table COURSE is 3NF.

COURSE (Course_Id, Course_Name, Course_Level, Course_Description, Category_id, Organization_Id)

Table Name: **MATERIAL**

I had already proven that the table MATERIAL is 3NF in the previous section.

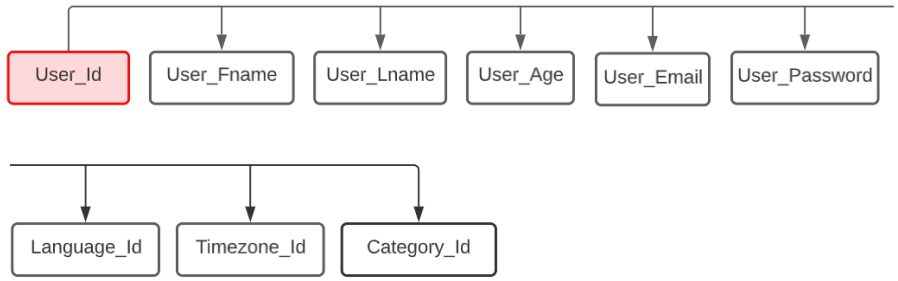


MATERIAL (Material_Id, Material_Name, Material_Desscription)

Table Name: **USERS**

	A	B	C	D	E	F	G	H	I	J	K
1	USER_ID	USER_FNAME	USER_LNAME	USER_AGE	USER_EMAIL	USER_PASSWORD	CATEGORY_ID	LANGUAGE_ID	TIMEZONE_ID		
2	154	Maurice	Daltrey	26	maurice.daltrey@teal.com	OFL9RFzYzSy.QtS	3	1	4		
3	155	Elizabeth	Brown	33	elizabeth.brown@thrasher.com	QGRxh/@+/MNvIT0	3	1	5		
4	156	Diane	Mason	40	diane.mason@trogon.com	z@iwCq9lWe79Yk3	4	1	2		
5	157	Dianne	Andrews	39	dianne.andrews@turnstone.com	IdYQK-QobdhY+/x	2	1	6		
6	158	Charles	Field	36	charles.field@becard.com	DS-pw1P9tTGi*3O	3	1	1		
7	159	Charles	Broderick	27	charles.broderick@bittern.com	pkq@pHn3PSwJe,H	2	1	3		
8	160	Isabella	Reed	25	isabella.reed@brant.com	kUF30q5Doft7Bjq	1	1	5		
9	161	Louis	Jackson	43	louis.jackson@caracara.com	7ScEuhGWJCXDxE?	3	1	1		
10	162	Louis	Edwards	39	louis.edwards@chachalaca.com	8L/Dh8FY3167o3t	2	1	3		
11	163	Doris	Dutt	38	doris.dutt@chukar.com	Sx2AH1O+i*AgomT	2	1	5		
12	164	Doris	Spacek	32	doris.spacek@flicker.com	WQXIAQB+lc3x4bD	2	1	5		
13	165	Kristin	Malden	26	kristin.malden@godwit.com	5LR@zpgRcsB,jYL	2	1	4		
14	166	Sissy	Puri	23	sissy.puri@grebe.com	LVdh9mLKEEs@SKn	2	1	6		
15	167	Doris	Bel Geddes	23	doris.belgeddes@grosbeak.com	nlurcyiNBf2*jJw	5	1	6		
16	168	Sissy	Warden	27	sissy.warden@jacana.com	M)X5w6tob6fEKxx	3	1	4		
17	169	Elia	Brando	33	elia.brando@junco.com	NUJ.pAMJI01otb7	2	1	6		
18	170	Mani	Fonda	27	mani.fonda@kinglet.com	hWEMyKLduJFbsOm	2	1	2		
19	171	Placido	Kubrick	27	placido.kubrick@scoter.com	3m*SMxVhl7tHi4K	4	1	1		
20	172	Claudia	Kurosawa	36	claudia.kurosawa@chukar.com	4JH7)dPfvF50wEk	2	1	5		
21	173	Maximilian	Henner	42	maximilian.henner@dunlin.com	)6Wr3*Ebg-gqQMx	2	1	4		
22	174	Sachin	Spielberg	36	sachin.spielberg@gadwall.com	*fGnyigEf-n1hZ,	5	1	1		
23	175	Sachin	Neeson	38	sachin.neeson@gallinule.com	EFFvg,ssi0XAl?2	1	1	2		
24	176	Sivaji	Landis	40	sivaji.landis@goldeneye.com	+4MsfP(Oi3NsUxv	3	1	1		
25	177	Mammutti	Pacino	38	mammutti.pacino@grebe.com	RqDEiG0eut9Z40S	3	1	6		
26	178	Elia	Fawcett	23	elia.fawcett@jacana.com	x1q3ZL.iSe0teG3	3	1	2		
27	179	Ishwarya	Roberts	31	ishwarya.roberts@lapwing.com	gWJH*l3G1cEn)O*	1	1	1		
28	180	Gustav	Steenburgen	31	gustav.steenburgen@pintail.com	lf+yPFmeHTc0dg2	2	1	5		
29	181	Markus	Rampling	44	markus.rampling@puffin.com	V@kcM?k(A/MhXIX	2	1	4		
30	182	Goldie	Slater	18	goldie.slater@pyrrhuloxia.com	mLr)tq7hsiCss(L	3	1	4		

Considering the above database, we can confirm that there is only one determinant (USER_ID), and there are no other dependencies. Thus, the table USERS is 3NF.

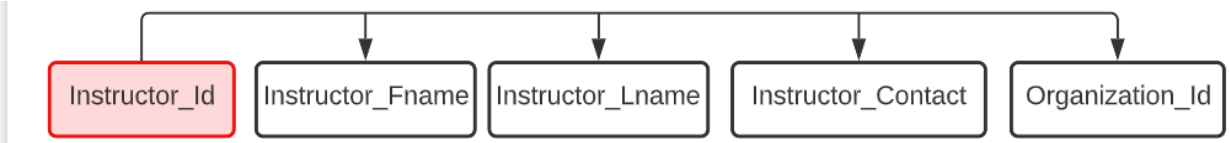


USERS (<u>User_Id</u>, User_Fname, User_Lname, User_Age, User_Email, User_Passoword, Language_Id, Timezone_Id, Category_Id)

Table Name: **INSTRUCTOR**

	A	B	C	D	E	F	G	H
1	INSTRUCTOR_ID	INSTRUCTOR_FNAME	INSTRUCTOR_LNAME	INSTRUCTOR_CONTACT	ORGANIZATION_ID			
2	1	Bryan	Dvrrie	bryan.dvrrie@redpoll.com	1			
3	2	Ajay	Sen	ajay.sen@trogon.com	1			
4	3	Carol	Jordan	carol.jordan@turnstone.com	1			
5	4	Carol	Bradford	carol.bradford@verdin.com	1			
6	5	Cary	Stockwell	cary.stockwell@vireo.com	1			
7	6	Cary	Olin	cary.olin@waterthrush.com	1			
8	7	Clara	Krige	clara.krige@whimbrel.com	1			
9	8	Clara	Ganesan	clara.ganesan@wigeon.com	1			
10	9	Ajay	Andrews	ajay.andrews@yellowthroat.com	2			
11	10	Kathy	Prashant	kathy.prashant@ani.com	2			
12	11	Graham	Neeson	graham.neeson@auklet.com	2			
13	12	Ian	Chapman	ian.chapman@avocet.com	2			
14	13	Danny	Wright	danny.wright@bittern.com	2			
15	14	Danny	Rourke	danny.rourke@brant.com	2			
16	15	Donald	Hunter	donald.hunter@chachalaca.com	2			
17	16	Graham	Spielberg	graham.spielberg@chukar.com	2			
18	17	Dan	Roberts	dan.roberts@nuthatch.com	3			
19	18	Edward	Oates	edward.oates@ovenbird.com	3			
20	19	Edward	Julius	edward.julius@parula.com	3			
21	20	Farrah	Quinlan	farrah.quinlan@phainopepla.com	3			
22	21	Farrah	Lange	farrah.lange@phalarope.com	3			
23	22	Hal	Stockwell	hal.stockwell@phoebe.com	3			
24	23	Malcolm	Kanth	malcolm.kanth@pipit.com	3			
25	24	Malcolm	Broderick	malcolm.broderick@plover.com	3			
26	25	Mary	Lemmon	mary.lemmon@puffin.com	4			
27	26	Mary	Collins	mary.collins@pyrrhuloxia.com	4			
28	27	Matt	Gueney	matt.gueney@redpoll.com	4			
29	28	Max	von Sydow	max.vonsydow@redstart.com	4			
30	29	Max	Schell	max.schell@sanderling.com	4			

Considering the above database, we can confirm that there is only one determinant (INSTRUCTOR_ID), and there are no other dependencies. Thus, the table INSTRUCTOR is 3NF.



INSTRUCTOR (Instructor_Id, Instructor_Fname, Instructor_Lname, Instructor_Contract, Organization_Id)

Table Name: **CATEGORY**

This table is composed of only two attributes: its id (primary key) and its name. Thus, obviously it's 3NF.

CATEGORY (Category_Id, Category_Name)

Table Name: **ORGANIZATION**

This table is composed of only two attributes: its id (primary key) and its name. Thus, obviously it's 3NF.

ORGANIZATION (Organization_Id, Organizaiton_Name)

Table Name: **LANGUAGE**

This table is composed of only two attributes: its id (primary key) and its name. Thus, obviously it's 3NF.

LANGUAGE (Language_Id, Language_Name)

Table Name: **TIMEZONE**

This table is composed of only two attributes: its id (primary key) and its name. Thus, obviously it's 3NF.

TIMEZONE (Timezone_Id, Timezone_Name)

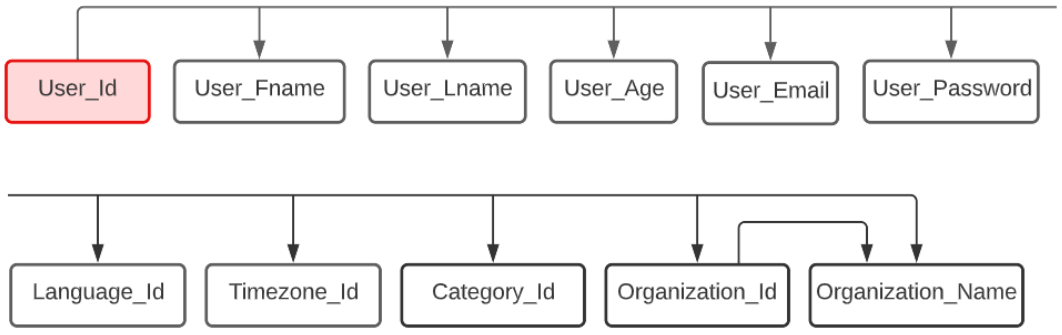
2. Add FDs and Change it to the 3rd Normal form

(1) USERS table: adding their school or company info

The following table and diagram are the new USERS database added their school or company information.

	A	B	C	D	E	F	G	H	I	J	K
1	USER_ID	USER_FNAME	USER_LNAME	USER_AGE	USER_EMAIL	USER_PASSWORD	CATEGORY_ID	LANGUAGE_ID	TIMEZONE_ID	ORGANIZATION_ID	ORGANIZATION_NAME
2	1	Kyle	Schneider	27	kyle.schneider	gEWLVSYcfWg2.?9	2	1	4	3	Microsoft
3	2	Kyle	Martin	18	kyle.martin@	cWnZcL2+5w/szwS	4	1	2	12	Durham College
4	4	Shelley	Peckinpah	38	shelley.peckin	uugjxAHZoeh(,Jq	1	1	3	8	Seneca College
5	5	Prem	Garcia	32	prem.garcia@	q/+@s1yWGjX2Q3	2	1	1	9	Humber College
6	6	Bo	Hitchcock	41	bo.hitchcock@	DgJv?p/p95xVvzn	3	1	2	6	Stanford University
7	7	Bob	McCarthy	25	bob.mccarthy	6ElB+(pUme@YtbS	5	1	2	9	Humber College
8	8	Dom	McQueen	28	dom.mcqueer	ciJMX7w5gLfNBzB	3	1	2	12	Durham College
9	10	Don	Siegel	37	don.siegel@bi	)Z-v3H.)k48?Bxr	1	1	4	12	Durham College
10	12	Shammi	Pacino	31	shammi.pacin	E+CSi7APP3bXyXN	2	2	2	12	Durham College
11	14	Sharmila	Fonda	34	sharmila.fond	piN@A6WPVzGYuI	4	2	2	10	George Brown College
12	15	Shelley	Taylor	31	shelley.taylor(	YLEp6r?)9.UIiS,	1	2	6	9	Humber College
13	16	Shyam	Plummer	29	shyam.plumm	S5?ctk4ZlEOk+@*	3	2	6	8	Seneca College
14	17	Silk	Kurosawa	24	silk.kurosawa(	273bpo1@fd7G2M	1	2	3	9	Humber College
15	18	Sivaji	Gielgud	21	sivaji.gielgud@	n.FwcN1Y-zqpbe6	3	2	3	5	University of Michigan
16	19	M. Emmet	Stockwell	34	m.emmet.stoc	72ZL/2I4XbzMn5d	2	3	3	2	IBM
17	20	M. Emmet	Olin	38	m.emmet.olin	uC52Jixm,vm.KoP	3	3	2	8	Seneca College
18	22	Mammutti	Sutherland	26	mammutti.sut	i4+6BGiVJkY1QsT	4	3	5	9	Humber College
19	23	Mani	Kazan	41	mani.kazan@	LYwlEDrQ9A-92jn	4	3	6	11	Centennial College
20	24	Mani	Buckley	35	mani.buckley(	9UVsc.JuTP8tMq*	4	3	7	8	Seneca College
21	25	Margaret	Ustinov	41	margaret.ustir	.UECb*p(,TzFCz@	4	3	4	8	Seneca College
22	26	Margaux	Krige	25	margaux.krige	WeHDRdOVcRVxd-	2	3	1	8	Seneca College
23	27	Margaux	Capshaw	24	margaux.caps	JgU5XJl-fh8auyn	3	3	6	8	Seneca College
24	28	Kevin	Goodman	20	kevin.goodma	nBxSw+LQQDi.rlE	4	3	7	10	George Brown College
25	29	Kevin	Cleveland	35	kevin.clevelan	5XWpwfK.KDLUnY/	4	3	1	11	Centennial College

Table Name: **USERS**
(User_Id, User_Fname, User_Lname, User_Age, User_Email, User_Password, Language_Id, Timezone_Id, Category_Id, Organization_Id, Organization_Name)



The above diagram is the 2nd dependency diagram. It is because organization_id determines organization_name It means that the diagram(table) has a transitive dependency (Organization_Id --> Organization_Name).

In order to convert it into the 3NF, we will split it into two tables: one is the USERS table, and the other is the ORGANIZATION table. However, in this case, I already have created the ORGANIZATION table. Thus, I can connect the USERS table and the existing ORGANIZATION table, rather than creating a new table to store the school or company that each user belongs to.

The converted diagram into 3NF is shown below (Note that the ORGANIZATION table has already been created in the previous section). Obviously, both two tables have no transitive dependencies as well as no partial dependencies.

Table Name: **USERS**
(User_Id, User_Fname, User_Lname, User_Age, User_Email, User_Password,
Language_Id, Timezone_Id, Category_Id, Organization_Id)

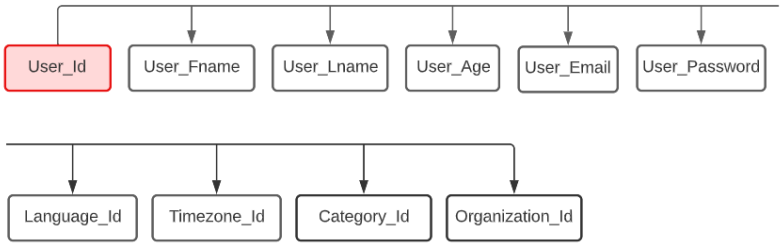
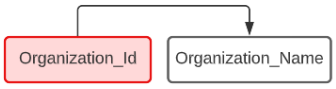


Table Name: **ORGANIZATION**
(Organization_Id, Organization_Name)



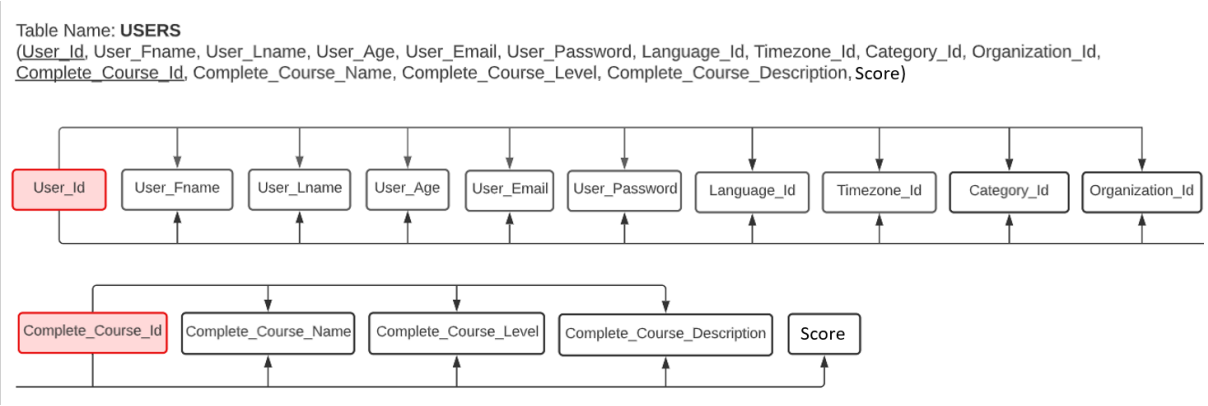
(2) USERS table: adding their completed course info and its score

The following table and diagram are the new USERS database added their completed course information and its score.

USER_ID	USER_FNAME	USER_LNAME	USER_AGE	USER_EMAIL	USER_PASSWORD	CATEGORY_ID	LANGUAGE_ID	TIMEZONE_ID	ORGANIZATION_ID	COURSE_ID	COURSE_NAME	COURSE_LEVEL	COURSE_DESCRIPTION	SCORE
1	Kyle	Schneider	27	kyle.schneide	gEWLVSYcfWg2.?9	2	1	4	3	13	Business Management	Intermediate	The course, Business Man	80
1	Kyle	Schneider	27	kyle.schneide	gEWLVSYcfWg2.?9	2	1	4	3	17	How to launch New Business	Advanced	The course, How to launcl	90
2	Kyle	Martin	18	kyle.martin@	cWnZcL2+5w/szwS	4	1	2	12	26	Introduction to Statistics	Beginner	The course, Introduction t	97
3	Meg	Derek	33	meg.derek@f	,GGPt3+Ktq*TqEE	4	1	6		20	Mathematics for Machine Learning	Intermediate	The course, Mathematics	91
3	Meg	Derek	33	meg.derek@f	,GGPt3+Ktq*TqEE	4	1	6		28	Mathematical Biostatistics	Advanced	The course, Mathematical	94
4	Shelley	Peckinpah	38	shelley.peckir	uugjxAHZoeh(,Jq	1	1	3	8	21	Introduction to Data Science	Beginner	The course, Introduction t	88
5	Prem	Garcia	32	prem.garcia@	q/+@s1yWGjX2Q3	2	1	1	9	17	How to launch New Business	Advanced	The course, How to launcl	96
6	Bo	Hitchcock	41	bo.hitchcock(	DgJv?p/p95xVvzn	3	1	2	6	23	How to create Algorithm in Python	Intermediate	The course, How to create	80
7	Bob	McCarthy	25	bob.mccarthy	6EIB+(pUme@Ytb5	5	1	2	9	35	Mental Health Risks of SNS	Beginner	The course, Mental Health	97
8	Dom	McQueen	28	dom.mcqueei	ciJMX7w5gLfNBzB	3	1	2	12	5	Introduction to Programming Languages	Beginner	The course, Introduction t	80
8	Dom	McQueen	28	dom.mcqueei	ciJMX7w5gLfNBzB	3	1	2	12	18	Coding: C and C++	Intermediate	The course, Coding: C and	87
9	Dom	Hoskins	24	dom.hoskins(	QMAH4ekUA*KA)(	1	1	2		2	Data Science on Google Cloud	Advanced	The course, Data Science (	86
10	Don	Siegel	37	don.siegel@b	)Z-v3H.)k48?Bxr	1	1	4	12	6	IBM Data Science	Beginner	The course, IBM Data Scie	83
11	Scott	Jordan	29	scott.jordan@	qRNO51Qo@tFqYr	1	2	3		6	IBM Data Science	Beginner	The course, IBM Data Scie	98
11	Scott	Jordan	29	scott.jordan@	qRNO51Qo@tFqYr	1	2	3		12	How to use Excel for Data Analysis	Beginner	The course, How to use Ex	99
12	Shammi	Pacino	31	shammi.pacir	E+CSi7APP3bXyXN	2	2	2	12	8	Introduction to Business	Beginner	The course, Introduction t	86
12	Shammi	Pacino	31	shammi.pacir	E+CSi7APP3bXyXN	2	2	2	12	14	Business Analytics	Advanced	The course, Business Anal	92
13	Sharmila	Kazan	32	sharmila.kaza	*uq8hjmSxOWxyb	1	2	1		6	IBM Data Science	Beginner	The course, IBM Data Scie	92
13	Sharmila	Kazan	32	sharmila.kaza	*uq8hjmSxOWxyb	1	2	1		21	Introduction to Data Science	Beginner	The course, Introduction t	92
14	Sharmila	Fonda	34	sharmila.fond	piN@A6WPVzGYu	4	2	2	10	24	History of Mathematics	Beginner	The course, History of Ma	72
15	Shelley	Taylor	31	shelley.taylor	YLEp6r?)9.UIIS,	1	2	6	9	2	Data Science on Google Cloud	Advanced	The course, Data Science (	90
16	Shyam	Plummer	29	shyam.plumn	S5?ctk4ZlEOk+@*	3	2	6	8	23	How to create Algorithm in Python	Intermediate	The course, How to create	97
17	Silk	Kurosawa	24	silk.kurosawa	273bpo1@fd7G2N	1	2	3	9	16	Machine Learning Methods	Intermediate	The course, Machine Lear	94
18	Sivaji	Gielgud	21	sivaji.gielgud(	n.FwcN1Y-zqpbe6	3	2	3	5	15	Introduction to Computer Science and Programming	Beginner	The course, Introduction t	85
18	Sivaji	Gielgud	21	sivaji.gielgud(	n.FwcN1Y-zqpbe6	3	2	3	5	23	How to create Algorithm in Python	Intermediate	The course, How to create	80
19	M. Emmet	Stockwell	34	m.emmet.sto	72ZL/2I4XbzMn5d	2	3	3	2	4	Business Strategies	Advanced	The course, Business Strat	98
20	M. Emmet	Olin	38	m.emmet.olir	uC52Jixm,vm.KoP	3	3	2	8	10	Computer Science	Intermediate	The course, Computer Sci	71
20	M. Emmet	Olin	38	m.emmet.olir	uC52Jixm,vm.KoP	3	3	2	8	18	Coding: C and C++	Intermediate	The course, Coding: C and	70
21	Malcolm	Field	21	malcolm.field	m1OzRYZPP-}J7ud	3	3	4		5	Introduction to Programming Languages	Beginner	The course, Introduction t	95
22	Mammutti	Sutherland	26	mammutti.su	i4+6BGiVJkY1QsT	4	3	5	9	28	Mathematical Biostatistics	Advanced	The course, Mathematical	73
23	Mani	Kazan	41	mani.kazan@	LYwlEDrQ9A-92jn	4	3	6	11	29	Discrete Mathematics	Beginner	The course, Discrete Math	79
24	Mani	Buckley	35	mani.buckley	9UJVsc,luTP8tMn*	4	3	7	8	29	Discrete Mathematics	Beginner	The course, Discrete Math	72

Given this database, the new USERS table will be converted to 1NF (the previous USERS table was 3NF). The reason is that partial dependencies exist in that table. In this table, primary keys are defined by two pairs, USER_ID and COURSE_ID. It is because the column “SCORE” depended on both attributes, not

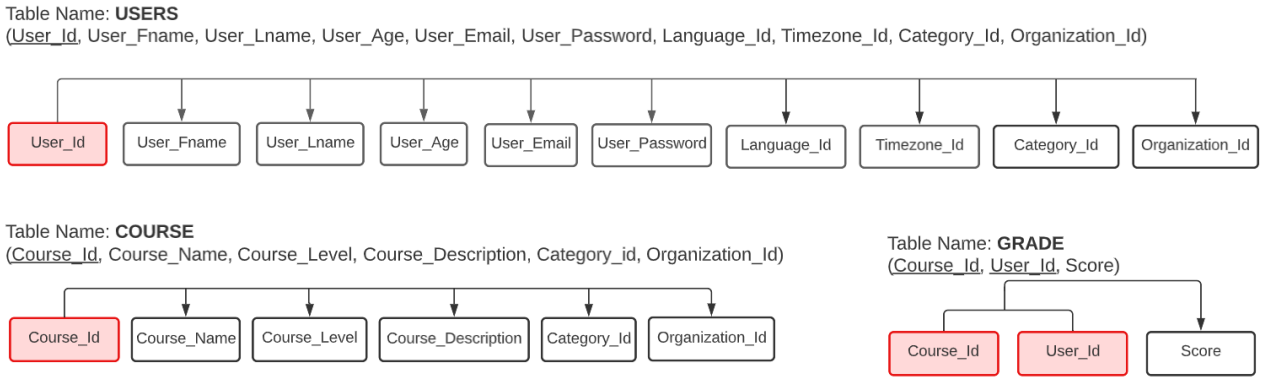
either one. Thus, considering the existence of multiple primary keys, the new USERS table has partial dependencies. For example, a part of the primary key (COURSE_ID) determines its name, level, and description. On the other hand, the table does not have transitive dependencies. It is because no other attributes except primary keys become determinants. The following diagram shows us a clearer structure of the USERS table about normalization form. (Note: In order to classify the completed course information, I added “Completed_” statement before each name about the completed course information tentatively.)



Partial dependencies:
- (User_Id --> User_Fname, User_Lname, User_Age, User_Email, User_Password, Language_Id, Timezone_Id, Category_Id, Organization_Id)

- (Complete_Course_Id --> Complete_Course_Name, Complete_Course_Level, Complete_Course_Description)

In order to convert 1NF to 2NF (in this case, 3NF due to no transitive dependencies), we will create new tables about the completed course information. The E-learning database already has the COURSE entity, so I decided to reuse it. Thus, I will not create a new table like COMPLETE_COURSE. However, I must create a table (defined GRADE) to store the users' scores for each course. Therefore, the 1NF diagram can be changed to the following.



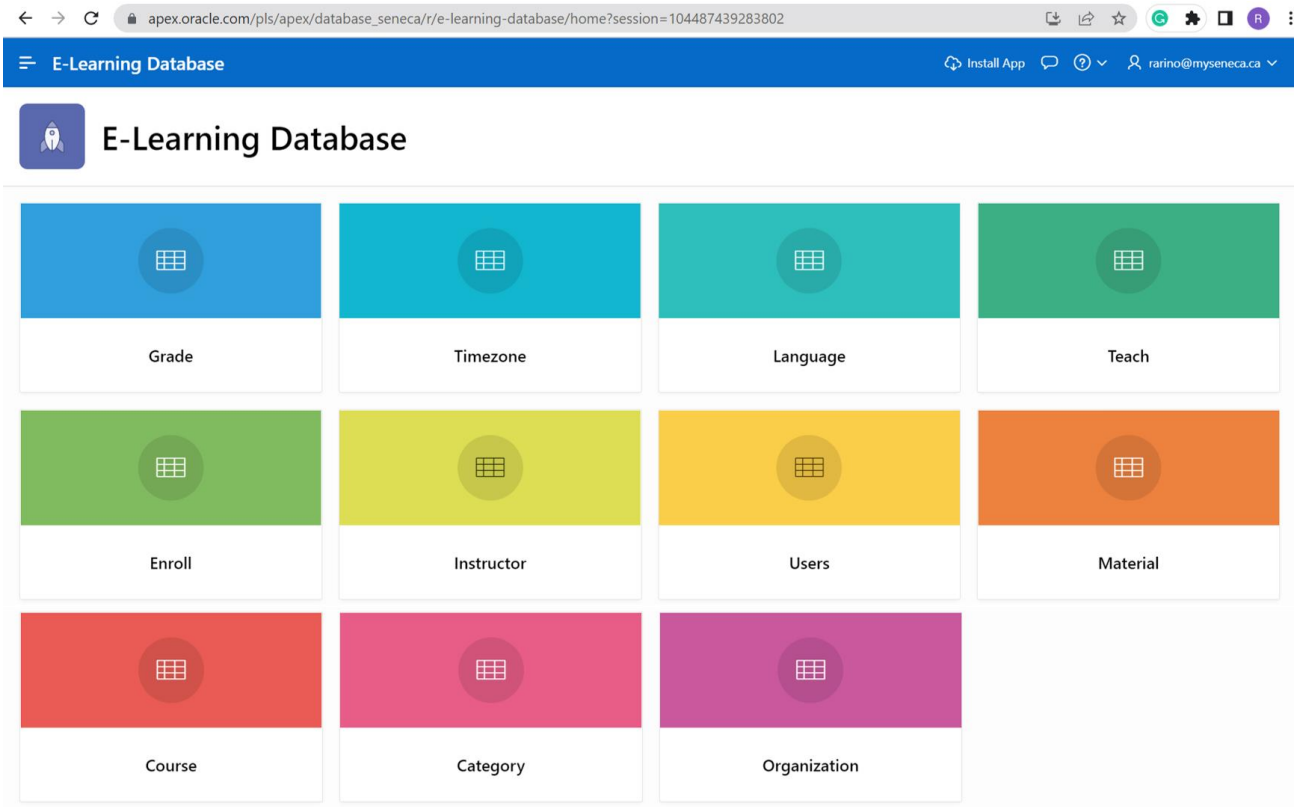
As I mentioned, this diagram doesn’t have transitive dependencies. Thus, the new USERS table with completed course info can be converted to 3NF.

3. Demonstration of GUI using Oracle APEX

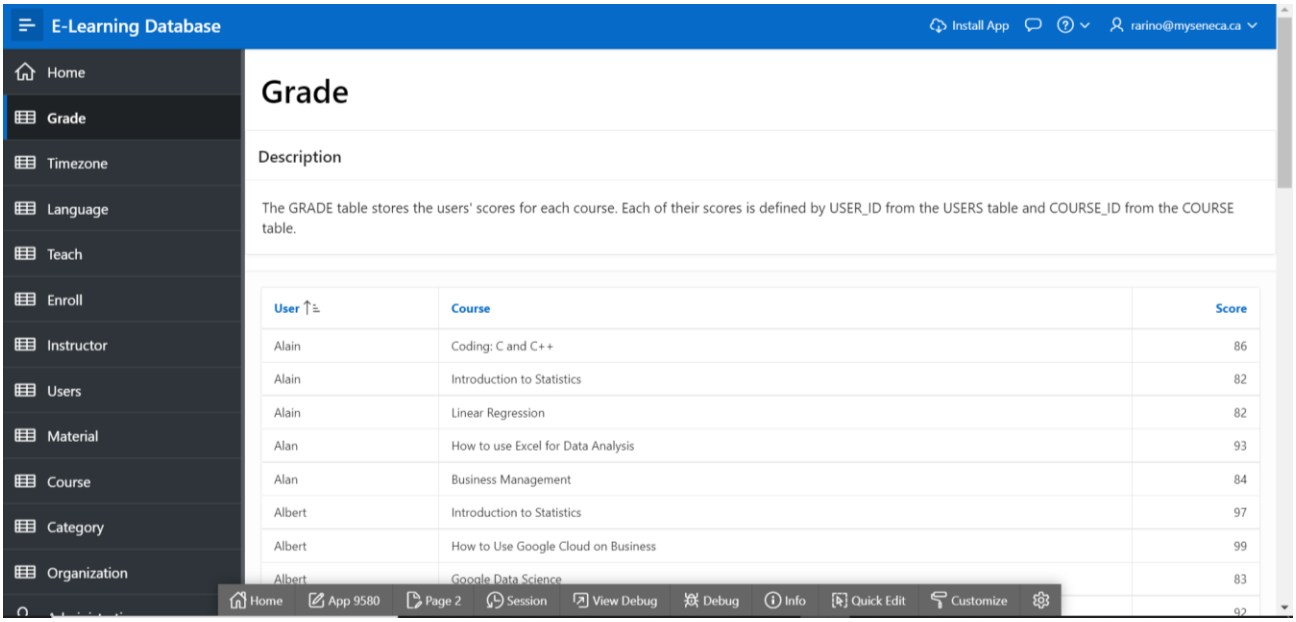
My database application link:
https://apex.oracle.com/pls/apex/database_seneca/r/e-learning-database/home?session=104487439283802

- Username: rarino@myseneca.ca
- Password: Arashi_Seneca

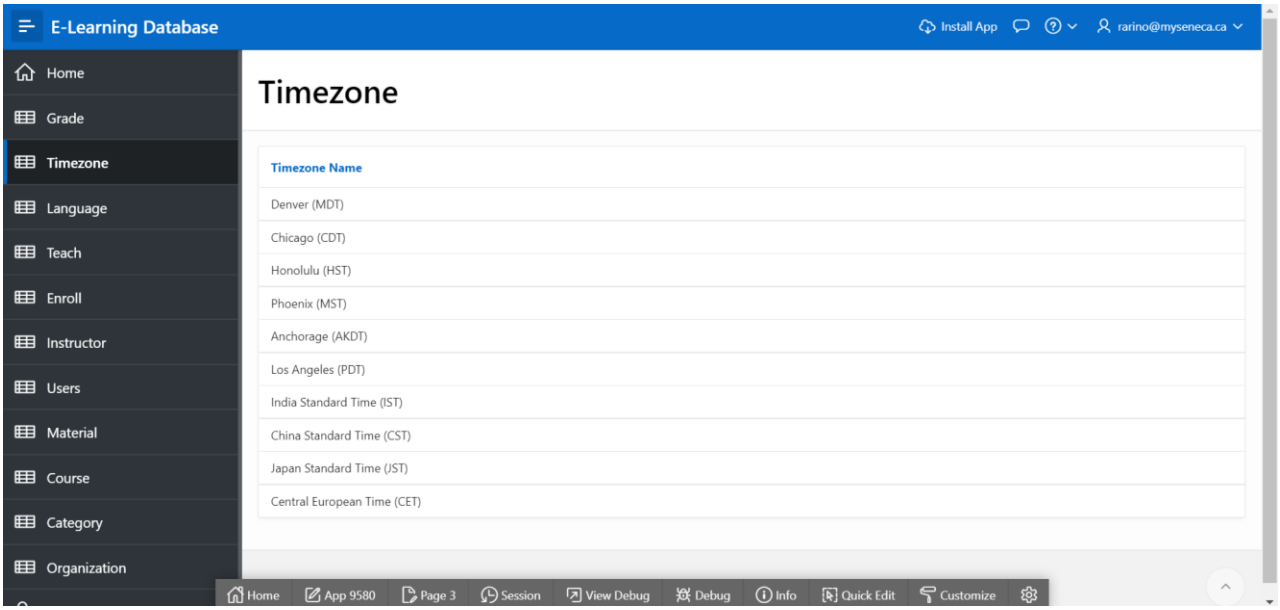
Home Page:



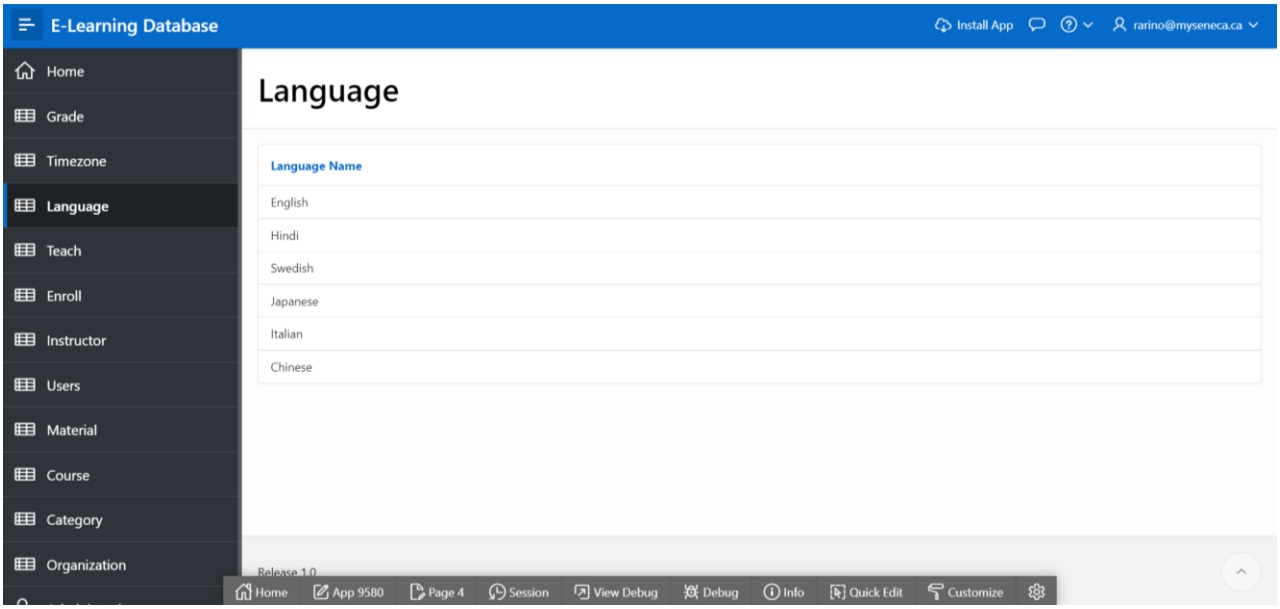
Grade table:



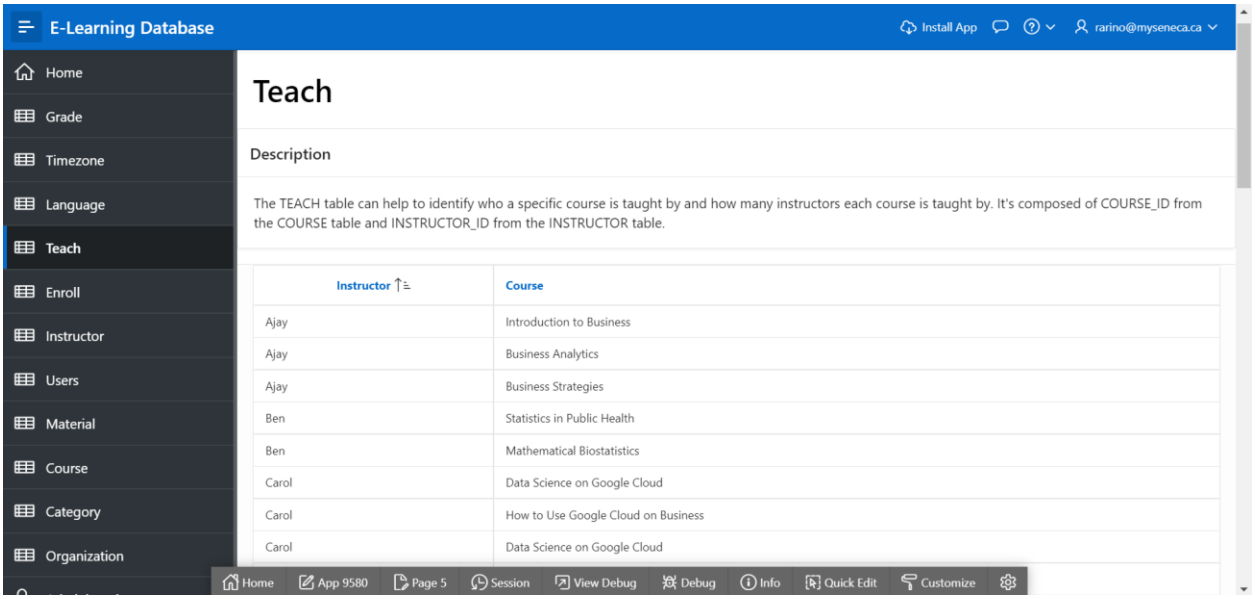
Timezone Table:



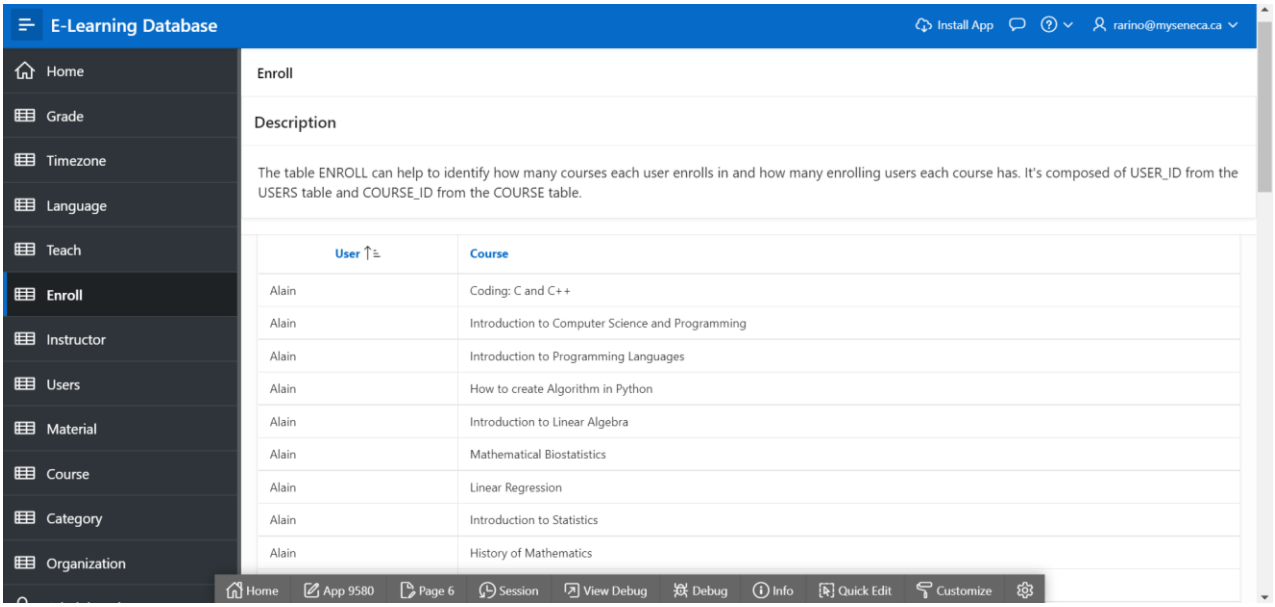
Language Table:



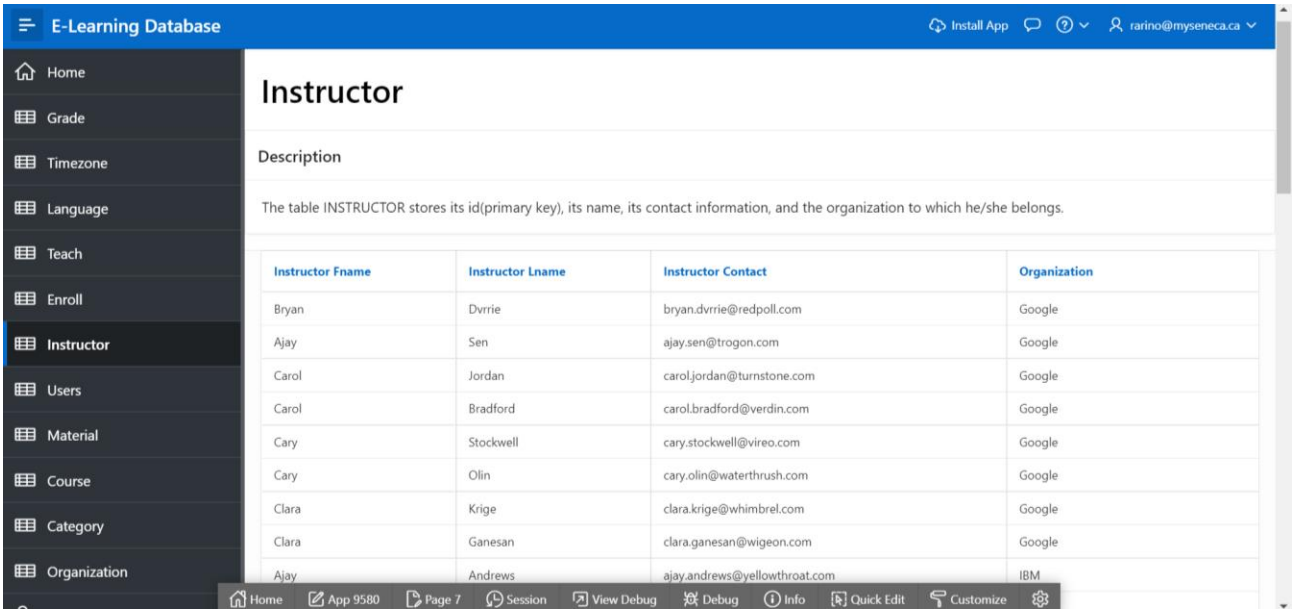
Teach Table:



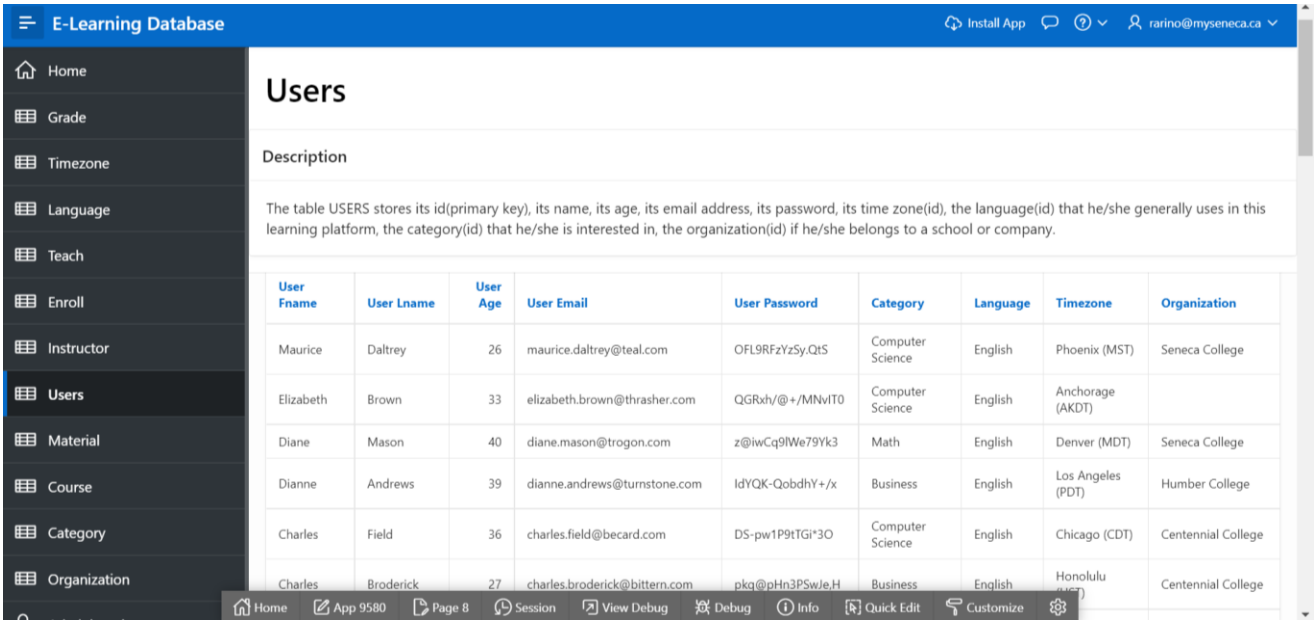
Enroll Table:



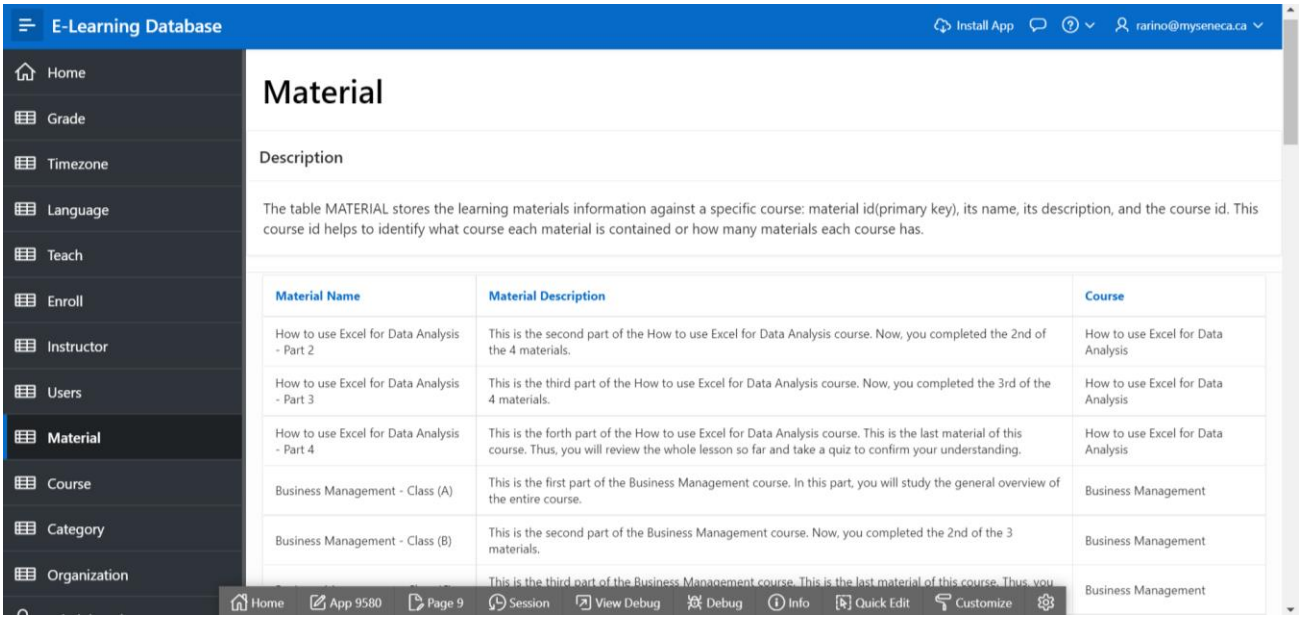
Instructor Table:



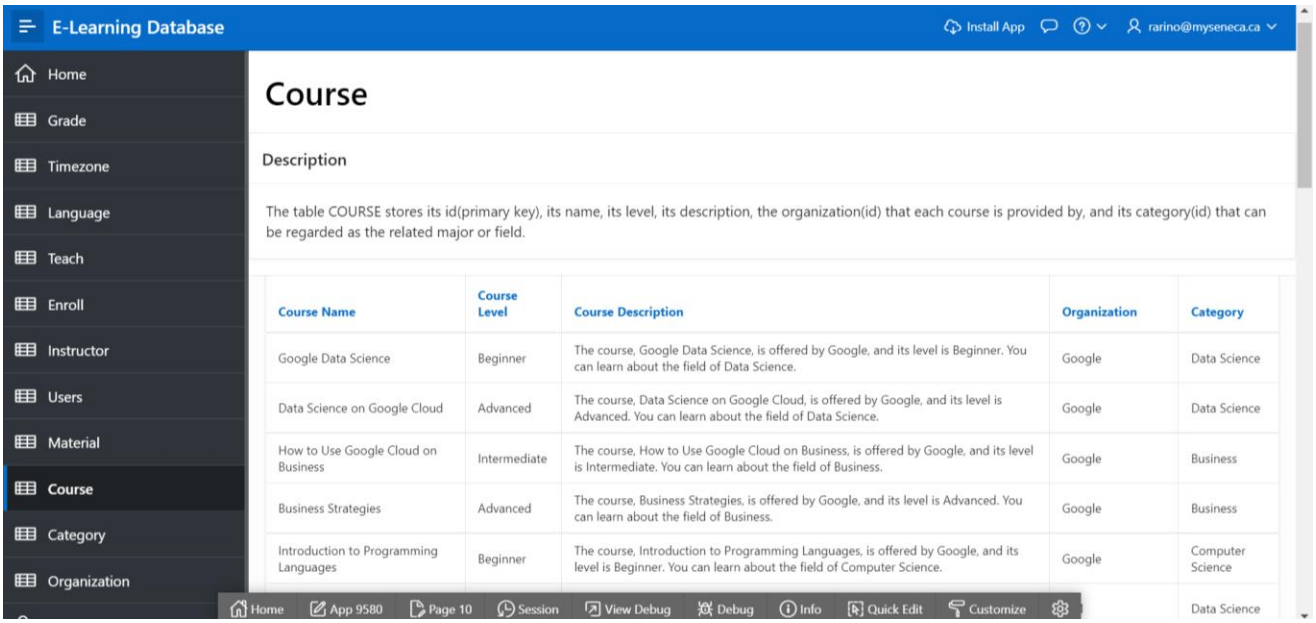
Users Table:



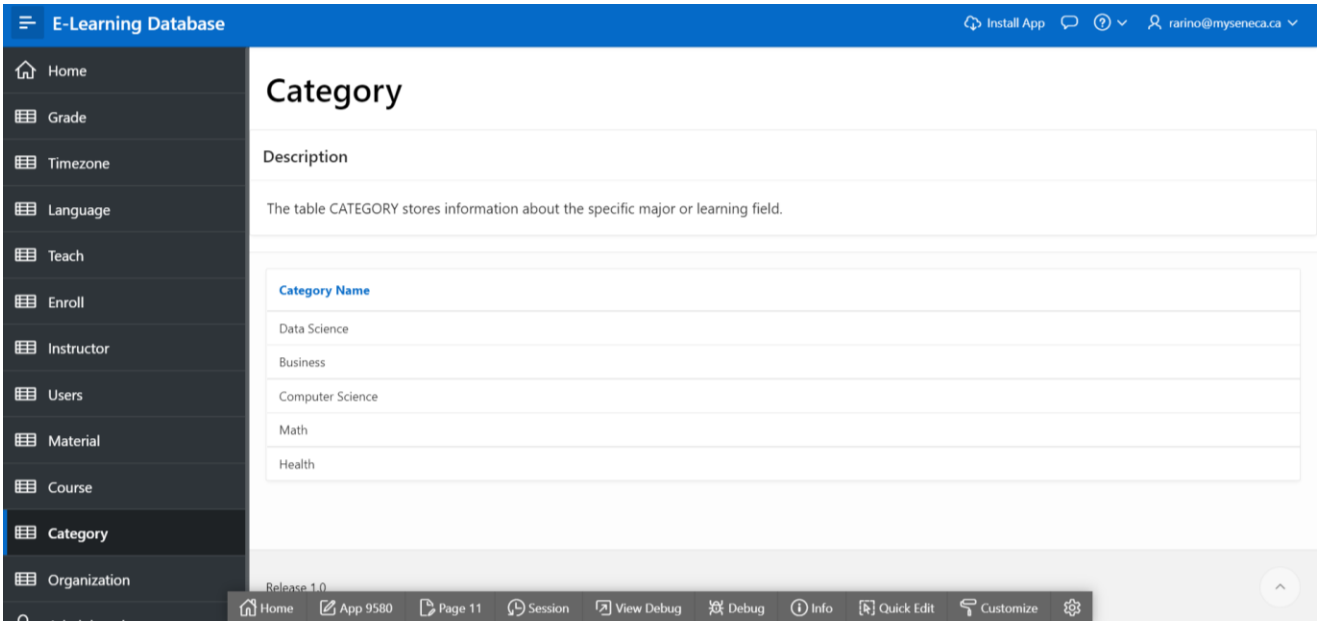
Material Table:



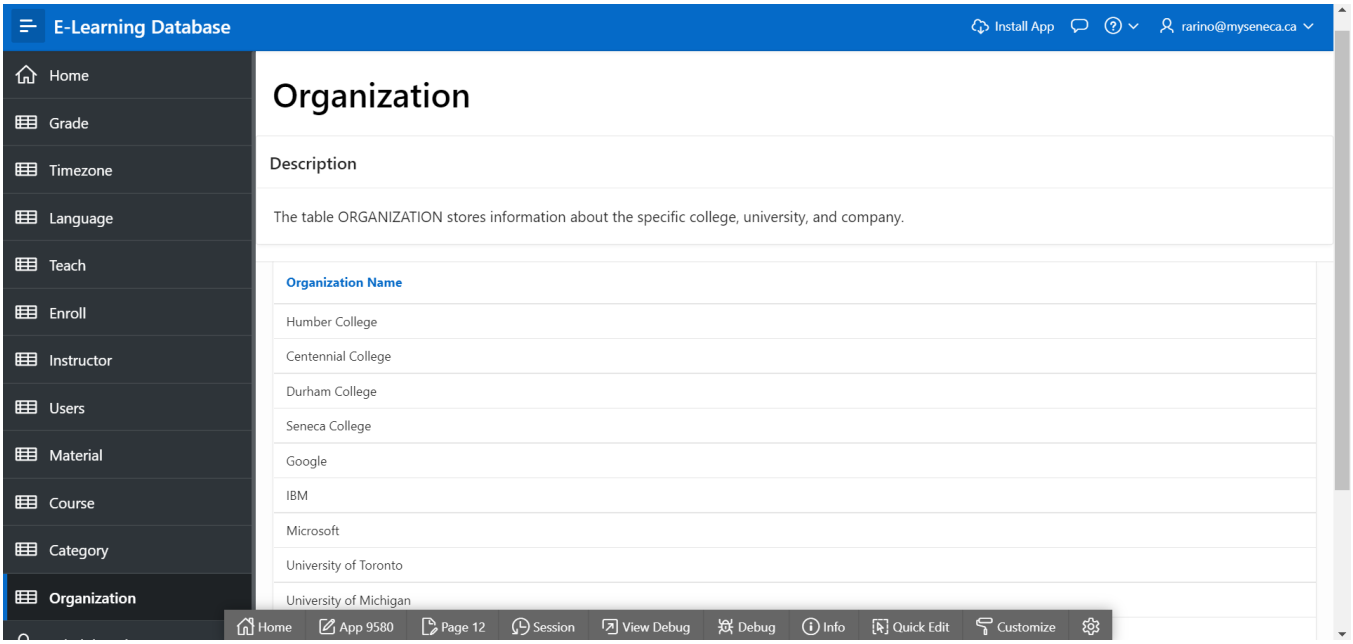
Course Table:



Category Table:

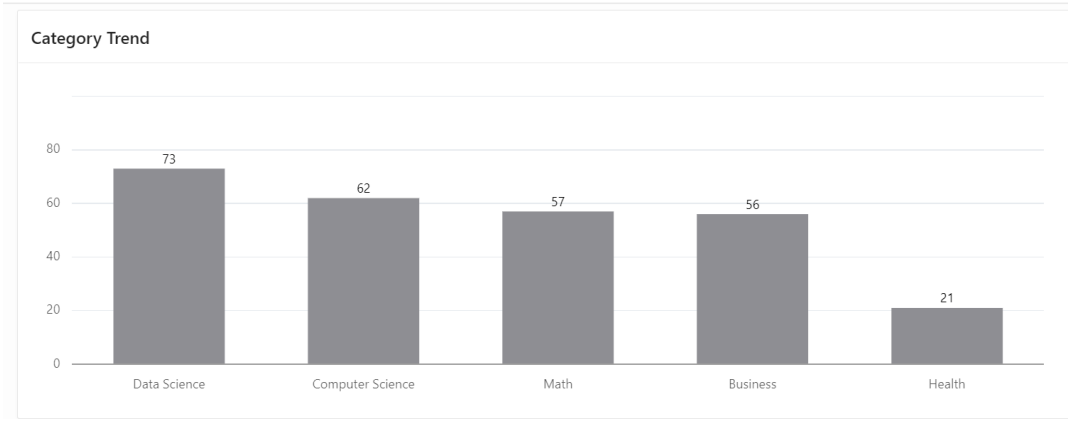


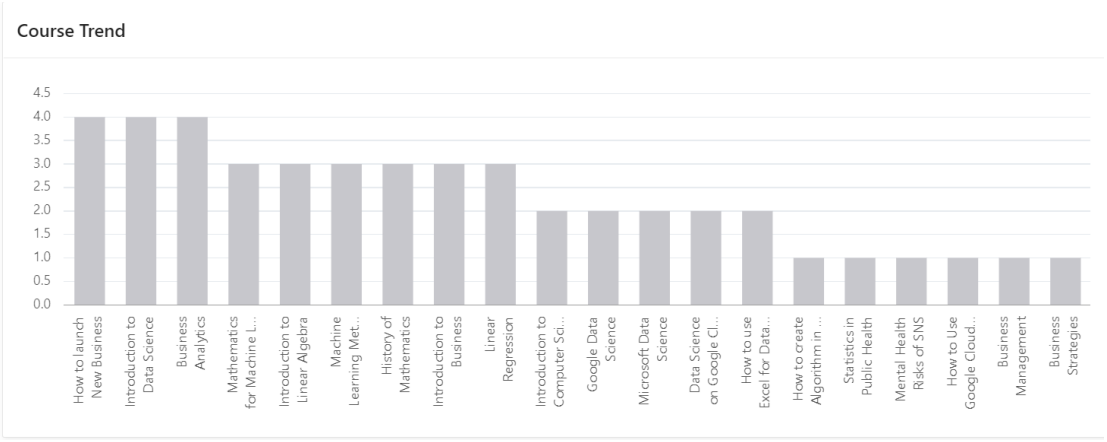
Organization Table:



Note: Each table's primary key is hidden. This is the default setting of Oracle APEX.

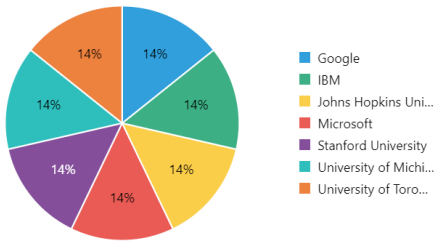
Statistics (Dashboard):





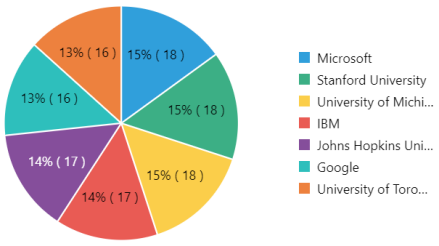
Organization Contribution - Courses

How many courses does each organization provide on this learning platform?

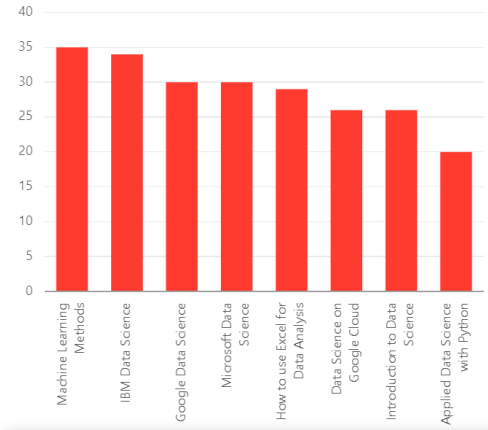


Organization Contribution - Materials

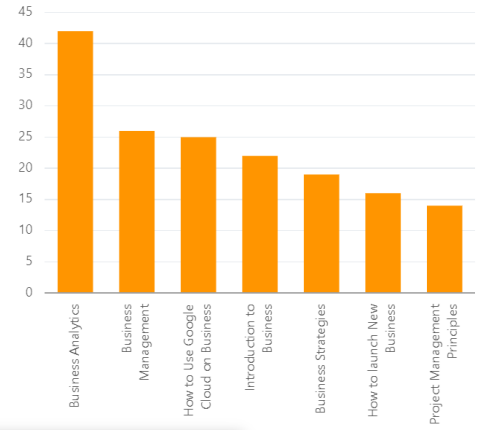
How many course materials does each organization provide on this learning platform?



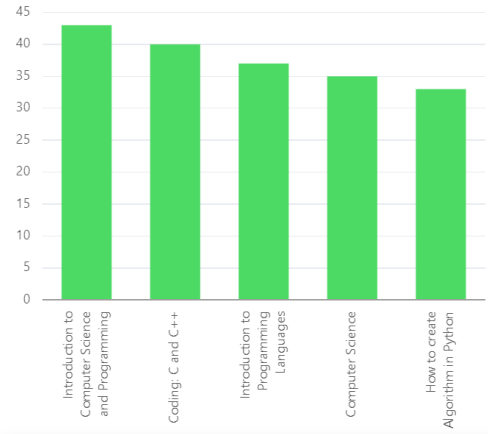
Data Science - Course Trend



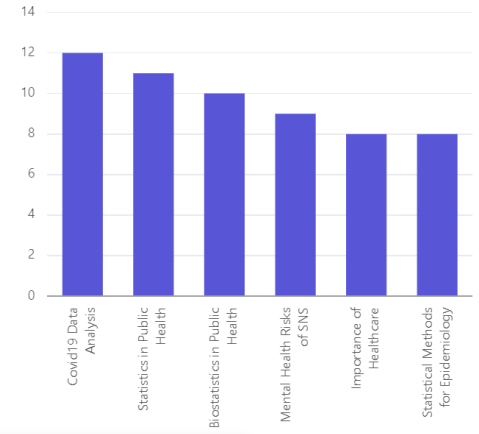
Computer Science - Course Trend



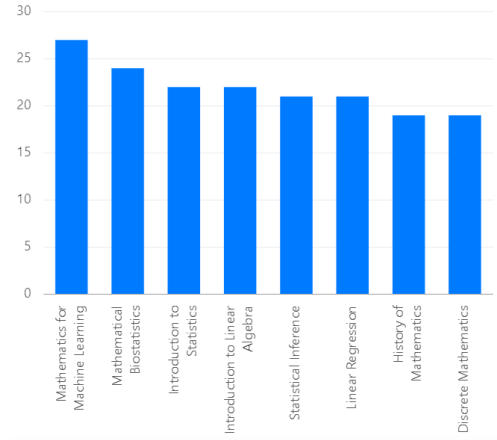
Math - Course Trend



Business- Course Trend



Health- Course Trend



4. Demonstration of GUI using Oracle APEX

Proof 3NF:

I have already proven that my database is 3NF concurrently in section 2 and 3.

SQL code:

Attached to it in a different file.

Design Experiences:

Through the BDD200 project 1 and 2, I created the E-learning database (model: Coursera). The core entity of this database is the COURSE entity. This entity is connected to all other entities directly and indirectly. More specifically, the COURSE has a weak relationship with MATERIAL, CATEGORY, and ORGANIZATION. And the COURSE is connected USERS and INSTRUCTOR through ENROLL and TEACH, respectively. Also, the GRADE table has a strong relationship with COURSE and USERS.

Thus, the purpose of this database is to execute the analysis of the course(s). For example, people can analyze the users' tendencies who are interested in a specific category(field). Suppose 100 users are interested in the Data Analysis field and taking a couple of courses related to its field. These users' information may help to develop the recommendation system for other users. In other words, if new users are also interested in the Data Analysis, analysts will utilize the data that can be obtained from those 100 users. In that phase, the E-learning database will help to create the ideal table to analyze the users' tendencies or trends. It is because the COURSE entity relates to the USERS entity, and each user has their interest field if they have already registered it. Also, if we focus on a specific course, we can obtain how many and what kinds of users are enrolling in this course. Making use of this data can also help to analyze the latest trend in terms of age.

All in all, each entity of the E-Learning database connects to each other, so it will support sophisticated and complicated data analysis.