China's K12 Online Education Data Management System

--based on Cassandra

Part 1:

Project Plan

Topic:

China's K12 Online Education Data Management System

Data Model:

Columnar

Target Platform:

Cassandra (Datastax or Cosmos DB)

Objective/Scope:

1.Study China’s online education environment

2.Compare provided courses on different platforms(Such as EDU,Zuoyebang,TAL,Yuanfudao etc)

3.Study the statistics of people using online education

4.Analyze growth rate in different platforms

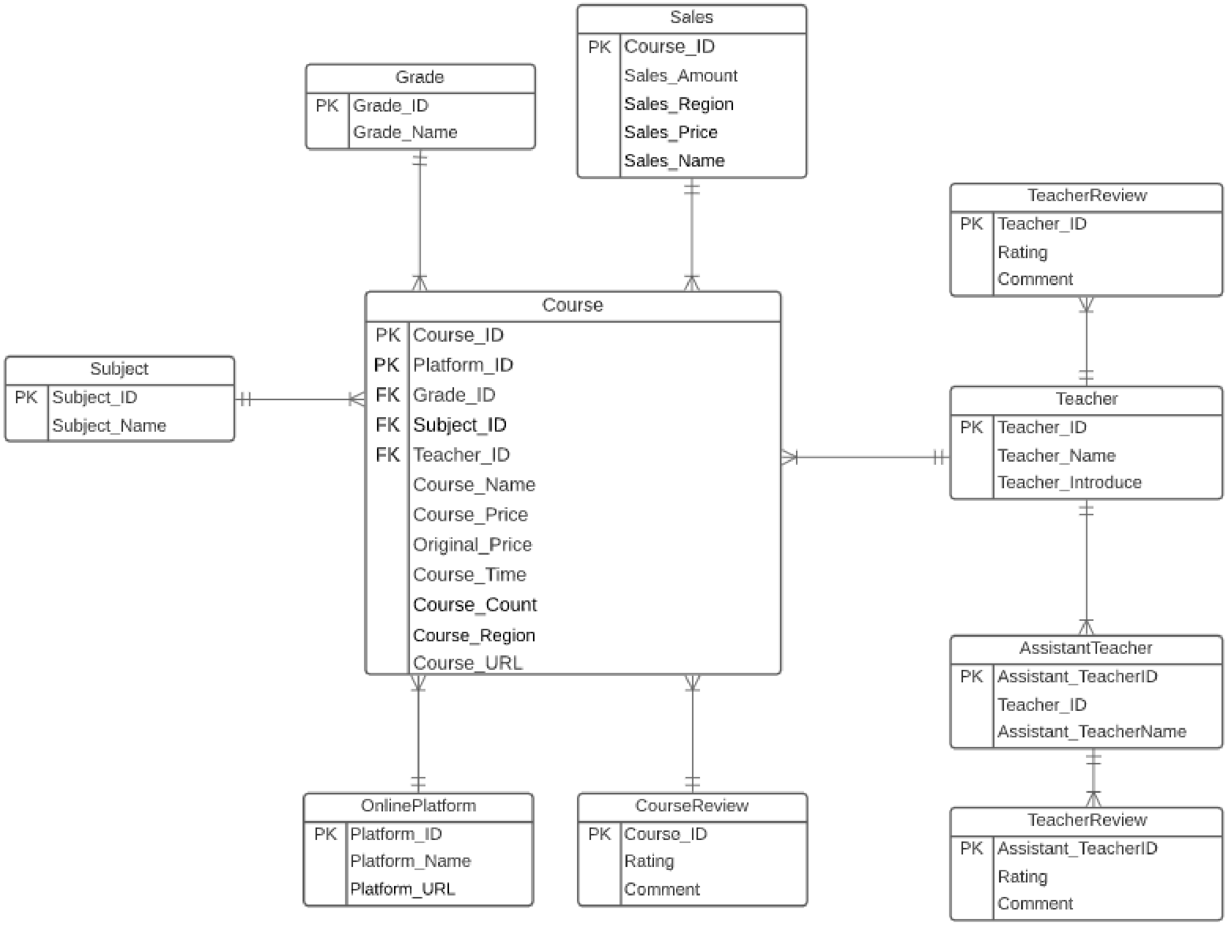
Visualizations Tool:

Tableau

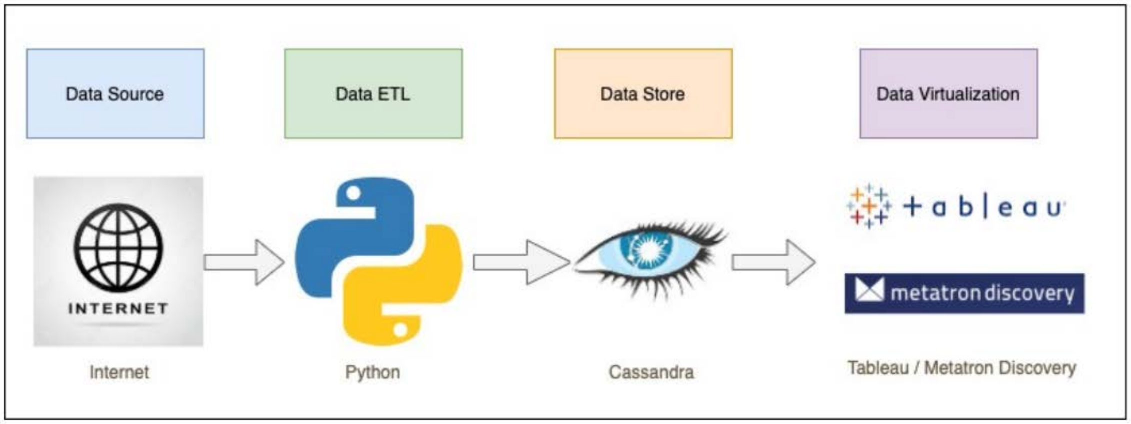
Part 2:

Design

ERD


Grade
PK Grade_ID
Grade_Name
Sales
PK Course_ID
Sales_Amount
Sales_Region
Sales_Price
Sales_Name
TeacherReview
PK Teacher_ID
Rating
Comment
Subject
PK Subject_ID
Subject_Name
Course
PK Course_ID
PK Platform_ID
FK Grade_ID
FK Subject_ID
FK Teacher_ID
Course_Name
Course_Price
Original_Price
Course_Time
Course_Count
Course_Region
Course_URL
Teacher
PK Teacher_ID
Teacher_Name
Teacher_Introduce
AssistantTeacher
PK Assistant_TeacherID
Teacher_ID
Assistant_TeacherName
OnlinePlatform
PK Platform_ID
Platform_Name
Platform_URL
CourseReview
PK Course_ID
Rating
Comment
TeacherReview
PK Assistant_TeacherID
Rating
Comment

Architecture Diagram

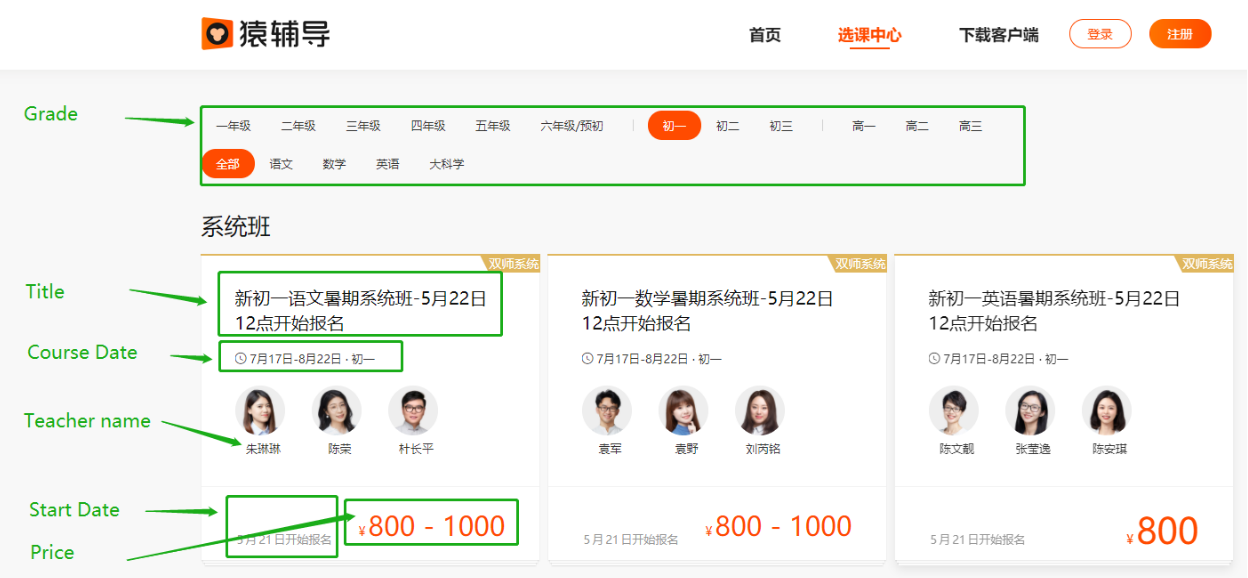


Part 3:

Data Implementation

1) Data Source：

https://www.yuanfudao.com/



系统班

新初一语文暑期系统班-5月22日12点开始报名

7月17日-8月22日 · 初一

朱琳琳　陈荣　村长平

5月21日开始报名　¥800 - 1000

新初一数学暑期系统班-5月22日12点开始报名

7月17日-8月22日 · 初一

袁军　袁野　刘芮铭

5月21日开始报名　¥800 - 1000

新初一英语暑期系统班-5月22日12点开始报名

7月17日-8月22日 · 初一

陈文靓　张莹逸　陈安琪

5月21日开始报名　¥800

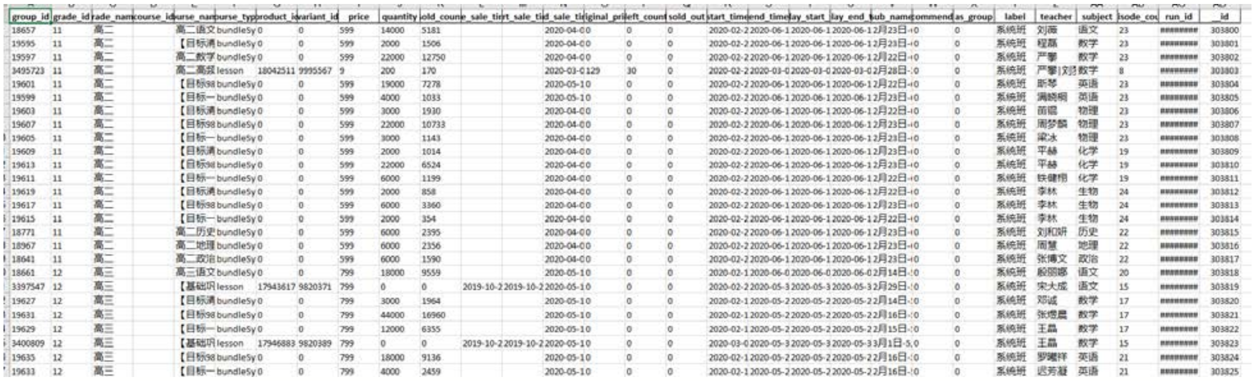
2) Web Scraping with Python:

```
import requests
import json
import xlwt
import pandas as pd

listxueduan=['xiaoxue','chuzhong','gaozhong']
listgrade=['1','2','3','4','5','6','7','8','9','10','11','12']
list_c_1=['2','3','201']
list_c_2=['1','2','3','4','5','6','7','8','14']
list_c_3=['1','2','3','4','5','6','7','8','9']

lesson = []
teacher = []


def get_key_value(info_dict,k,j):
    try:
        info_dict[k] = j[k]
    except KeyError as e:
        for j_key in j:
            if isinstance(j[j_key], dict):
                get_key_value(info_dict,k,j[j_key])
list=[]
list11=[]
list12=[]
list13=[]


urllist=[]
```

```
for i in listxueduan:
    if i=='xiaoxue':
        for j1 in range(1,7):
            #print(j1)
            g=j1
            for c1 in list_c_1:
                c=c1
                url = 'https://www.yuanfudao.com/tutor-student-
lesson/api/homepage?_productId=374&platform=www&version=5.11.0&UDID=4d27a58f757db800e
339317f1c245223&timestamp=1532658211041&startCursor=0&limit=18&grade='+str(g)+'&chann
elId='+str(c)+'&studyPhase='+str(i)+'&withNextGrade=false'
                urllist.append(url)
                list11.append(i)
                list12.append(j1)
                list13.append(c1)

    elif i=='chuzhong':
        for j2 in range(7,10):
            #print(j2)
            g=j2
            for c2 in list_c_2:
                c=c2
                url = 'https://www.yuanfudao.com/tutor-student-
lesson/api/homepage?_productId=374&platform=www&version=5.11.0&UDID=4d27a58f757db800e
339317f1c245223&timestamp=1532658211041&startCursor=0&limit=18&grade='+str(g)+'&chann
elId='+str(c)+'&studyPhase='+str(i)+'&withNextGrade=false'
                urllist.append(url)
                list11.append(i)
                list12.append(j2)
                list13.append(c2)
    else :
        for j3 in range(10,13):
            #print(j3)
            g=j3
            for c3 in list_c_3:
                c=c3
                url = 'https://www.yuanfudao.com/tutor-student-
lesson/api/homepage?_productId=374&platform=www&version=5.11.0&UDID=4d27a58f757db800e
339317f1c245223&timestamp=1532658211041&startCursor=0&limit=18&grade='+str(g)+'&chann
elId='+str(c)+'&studyPhase='+str(i)+'&withNextGrade=false'
                urllist.append(url)
                list11.append(i)
                list12.append(j2)
                list13.append(c3)
study_dict={
         'studyPhase':list11,
         'grade':list12,
         'channelId':list13
         }
def main():
    for i in range(0,72):
        url = urllist[i]
        s1=study_dict['studyPhase'][i]
        s2=study_dict['grade'][i]
        s3=study_dict['channelId'][i]
        r = requests.get(url)
        json_r= r.json()['list']
        for j in json_r:
            info_dict={
                    'studyPhase':s1,
                    'grade':s2,
                    'channelId':s3,
                    'id': None,
                    'minPrice': None,
                    'maxPrice': None,
                    'name': None,
                    'soldCount': None,
                    'price':None,
                    'subName': None,
                    'teachers':None
                    }
            for k in info_dict:
                get_key_value(info_dict,k,j)

            print(info_dict)
            lesson.append(info_dict)

            for t in j['teachers']:
                 teacher_dict={
                            'lessonid':info_dict['id'],
                            'id':t['id'],
                            'nickname':t['nickname'],
                            'avatar':t['avatar']
                                }
                 teacher.append(teacher_dict)



        work=xlwt.Workbook()
        sheet1=work.add_sheet('sheet1',cell_overwrite_ok=True)

head=['studyPhase','grade','channelId','id','minPrice','maxPrice','name','soldCount',
'price','subName']
        y=0
        for item in head:
            sheet1.write(0,y,item)
            y+=1
        x=1

        for item in lesson:
            if isinstance(item,dict):
                for head_item in head:
                    if head_item in item.keys():
                        y=head.index(head_item)
                        sheet1.write(x,y,item[head_item])
            x+=1
        work.save('yuanfudao new2.xls')

        work2=xlwt.Workbook()
        sheet1=work2.add_sheet('sheet1',cell_overwrite_ok=True)
        head2=['lessonid','id','nickname','avatar']
        y=0
        for item in head2:
            sheet1.write(0,y,item)
            y+=1
        x=1

        for item in teacher:
            if isinstance(item,dict):
                for head_item in head2:
                    if head_item in item.keys():
                        y=head2.index(head_item)
                        sheet1.write(x,y,item[head_item])
            x+=1
        work2.save('yuanfudao teacher.xls')



if __name__ == '__main__':
    main()
```

3) Rawdata

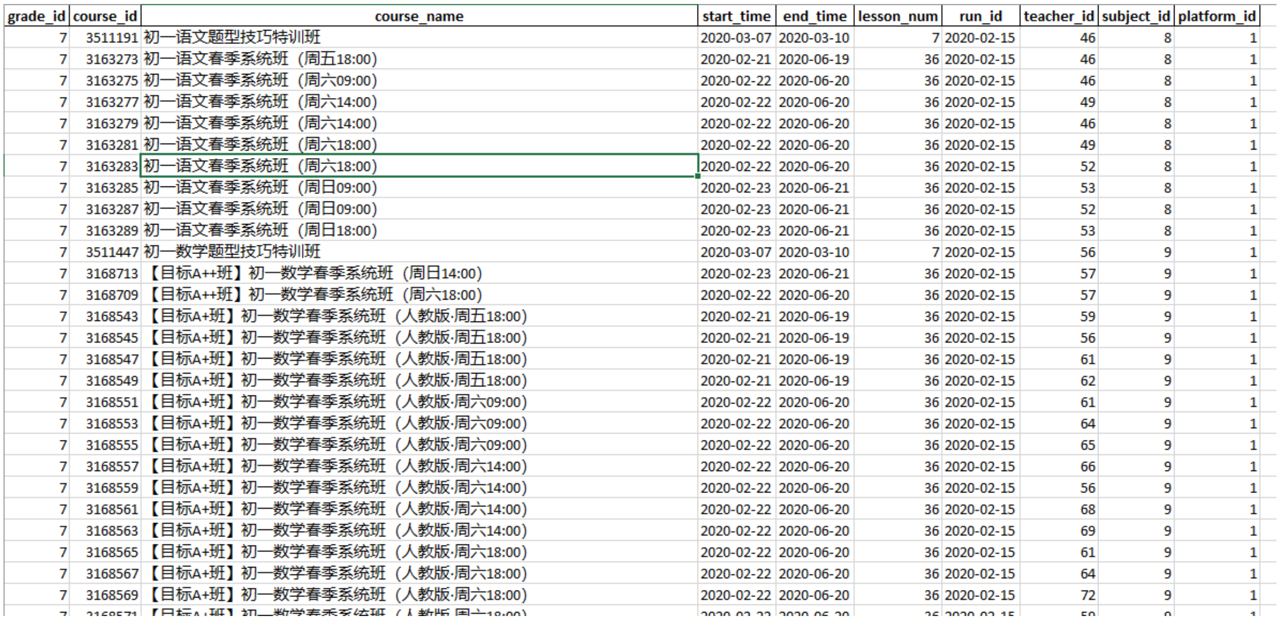
Data from (2020/01/01-2020/12/31,every month)



group_id	grade_id	rade_nam	course_id	urse_nan	ourse_typ	roduct_id	variant_id	price	quantity	old_coun	e_sale_tir	rt_sale_ti	d_sale_ti	iginal_pri	eft_count	sold_out	start_time	end_time	lay_start_	lay_end_	ub_name	commend	as_group	label	teacher	subject	isode_cou	run_id	_id
18657	11	高二		高二语文	bundleSy	0	0	599	14000	5181			2020-04-0	0	0	0	2020-02-2	2020-06-1	2020-06-1	2020-06-1	2月23日-	0	0	系统班	刘薇	语文	23	########	303800
19595	11	高二		【目标清	bundleSy	0	0	599	2000	1506			2020-04-0	0	0	0	2020-02-2	2020-06-1	2020-06-1	2020-06-1	2月23日-	0	0	系统班	程磊	数学	23	########	303801
19597	11	高二		高二数学	bundleSy	0	0	599	22000	12750			2020-04-0	0	0	0	2020-02-2	2020-06-1	2020-06-1	2020-06-1	2月22日-	0	0	系统班	严攀	数学	23	########	303802
3495723	11	高二		高二高频	lesson	18042511	9995567	9	200	170			2020-03-0	129	30	0	2020-02-2	2020-03-0	2020-03-0	2020-03-0	2月28日-	0	0	系统班	严攀|刘彭	数学	8	########	303803
19601	11	高二		【目标98	bundleSy	0	0	599	19000	7278			2020-05-1	0	0	0	2020-02-2	2020-06-1	2020-06-1	2020-06-1	2月22日-	0	0	系统班	靳琴	英语	23	########	303804
19599	11	高二		【目标一	bundleSy	0	0	599	4000	1033			2020-05-1	0	0	0	2020-02-2	2020-06-1	2020-06-1	2020-06-1	2月23日-	0	0	系统班	满晓桐	英语	23	########	303805
19603	11	高二		【目标清	bundleSy	0	0	599	3000	1930			2020-04-0	0	0	0	2020-02-2	2020-06-1	2020-06-1	2020-06-1	2月22日-	0	0	系统班	苗瑶	物理	23	########	303806
19607	11	高二		【目标98	bundleSy	0	0	599	22000	10733			2020-04-0	0	0	0	2020-02-2	2020-06-1	2020-06-1	2020-06-1	2月23日-	0	0	系统班	周梦麒	物理	23	########	303807
19605	11	高二		【目标一	bundleSy	0	0	599	3000	1143			2020-04-0	0	0	0	2020-02-2	2020-06-1	2020-06-1	2020-06-1	2月23日-	0	0	系统班	梁冰	物理	23	########	303808
19609	11	高二		【目标清	bundleSy	0	0	599	2000	1014			2020-04-0	0	0	0	2020-02-2	2020-06-1	2020-06-1	2020-06-1	2月23日-	0	0	系统班	平赫	化学	19	########	303809
19613	11	高二		【目标98	bundleSy	0	0	599	22000	6524			2020-04-0	0	0	0	2020-02-2	2020-06-1	2020-06-1	2020-06-1	2月23日-	0	0	系统班	平赫	化学	19	########	303810
19611	11	高二		【目标一	bundleSy	0	0	599	6000	1199			2020-04-0	0	0	0	2020-02-2	2020-06-1	2020-06-1	2020-06-1	2月22日-	0	0	系统班	铁健楠	化学	19	########	303811
19619	11	高二		【目标清	bundleSy	0	0	599	2000	858			2020-04-0	0	0	0	2020-02-2	2020-06-1	2020-06-1	2020-06-1	2月22日-	0	0	系统班	李林	生物	24	########	303812
19617	11	高二		【目标98	bundleSy	0	0	599	6000	3360			2020-04-0	0	0	0	2020-02-2	2020-06-1	2020-06-1	2020-06-1	2月23日-	0	0	系统班	李林	生物	24	########	303813
19615	11	高二		【目标一	bundleSy	0	0	599	2000	354			2020-04-0	0	0	0	2020-02-2	2020-06-1	2020-06-1	2020-06-1	2月22日-	0	0	系统班	李林	生物	24	########	303814
18771	11	高二		高二历史	bundleSy	0	0	599	6000	2395			2020-04-0	0	0	0	2020-02-2	2020-06-1	2020-06-1	2020-06-1	2月23日-	0	0	系统班	刘和妍	历史	22	########	303815
18967	11	高二		高二地理	bundleSy	0	0	599	6000	2356			2020-04-0	0	0	0	2020-02-2	2020-06-1	2020-06-1	2020-06-1	2月23日-	0	0	系统班	周慧	地理	22	########	303816
18641	11	高二		高二政治	bundleSy	0	0	599	6000	1590			2020-04-0	0	0	0	2020-02-2	2020-06-1	2020-06-1	2020-06-1	2月23日-	0	0	系统班	张博文	政治	22	########	303817
18661	12	高三		高三语文	bundleSy	0	0	799	18000	9559			2020-05-1	0	0	0	2020-02-1	2020-06-0	2020-06-0	2020-06-0	2月14日-	0	0	系统班	殷丽娜	语文	20	########	303818
3397547	12	高三		【基础巩	lesson	17943617	9820371	799	0	0	2019-10-2	2019-10-2	2020-05-1	0	0	0	2020-02-2	2020-05-3	2020-05-3	2020-05-3	2月29日-	0	0	系统班	宋大成	语文	15	########	303819
19627	12	高三		【目标清	bundleSy	0	0	799	3000	1964			2020-05-1	0	0	0	2020-02-1	2020-05-2	2020-05-2	2020-05-2	2月14日-	0	0	系统班	郑诚	数学	17	########	303820
19631	12	高三		【目标98	bundleSy	0	0	799	44000	16960			2020-05-1	0	0	0	2020-02-1	2020-05-2	2020-05-2	2020-05-2	2月16日-	0	0	系统班	张煜晨	数学	17	########	303821
19629	12	高三		【目标一	bundleSy	0	0	799	12000	6355			2020-05-1	0	0	0	2020-02-1	2020-05-2	2020-05-2	2020-05-2	2月15日-	0	0	系统班	王晶	数学	17	########	303822
3400809	12	高三		【基础巩	lesson	17946883	9820389	799	0	0	2019-10-2	2019-10-2	2020-05-1	0	0	0	2020-03-0	2020-05-3	2020-05-3	2020-05-3	3月1日-5,	0	0	系统班	王晶	数学	15	########	303823
19635	12	高三		【目标98	bundleSy	0	0	799	18000	9136			2020-05-1	0	0	0	2020-02-1	2020-05-2	2020-05-2	2020-05-2	2月16日-	0	0	系统班	罗曦祥	英语	21	########	303824
19633	12	高三		【目标一	bundleSy	0	0	799	4000	2459			2020-05-1	0	0	0	2020-02-1	2020-05-2	2020-05-2	2020-05-2	2月16日-	0	0	系统班	迟芳凝	英语	21	########	303825

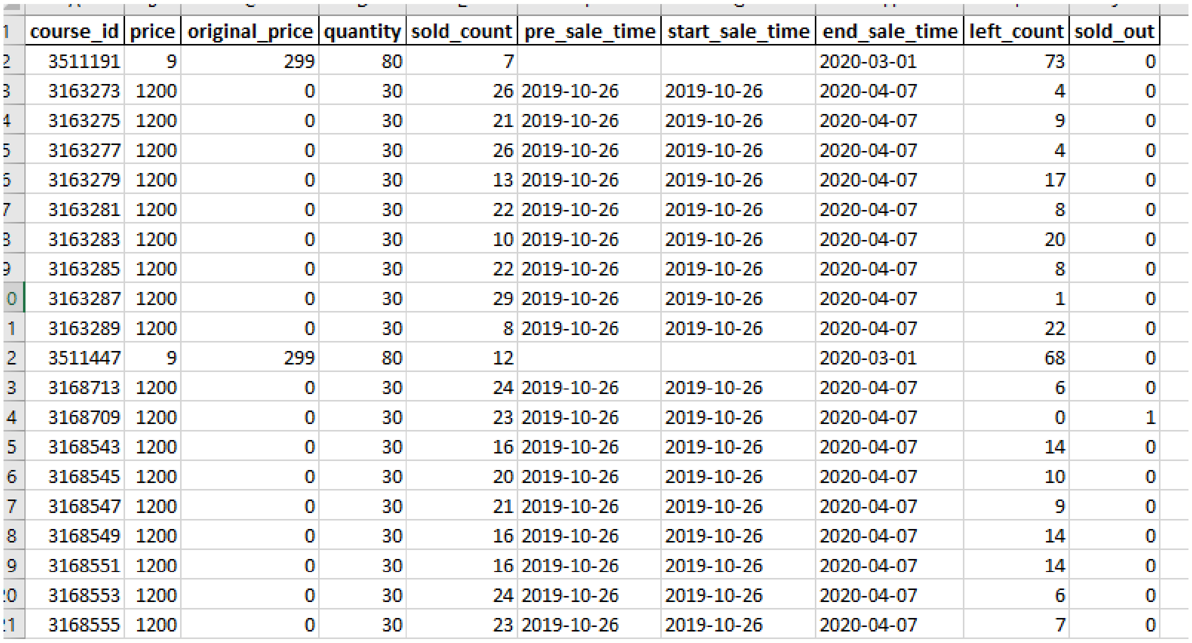
4) After data cleaning

a) course info table



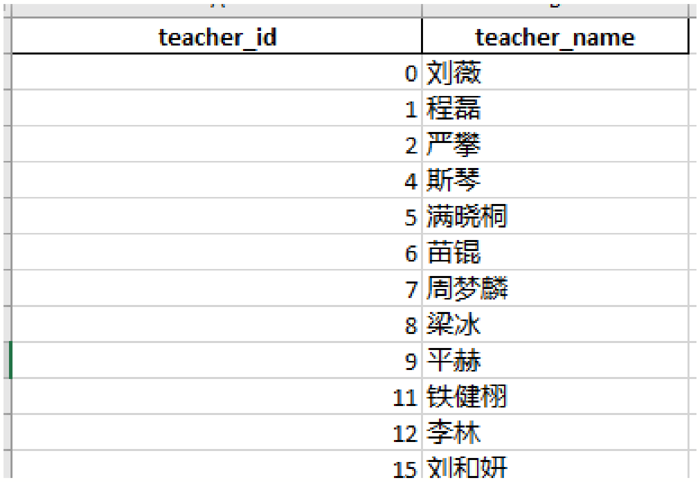
grade_id	course_id	course_name	start_time	end_time	lesson_num	run_id	teacher_id	subject_id	platform_id
7	3511191	初一语文题型技巧特训班	2020-03-07	2020-03-10	7	2020-02-15	46	8	1
7	3163273	初一语文春季系统班（周五18:00)	2020-02-21	2020-06-19	36	2020-02-15	46	8	1
7	3163275	初一语文春季系统班（周六09:00)	2020-02-22	2020-06-20	36	2020-02-15	46	8	1
7	3163277	初一语文春季系统班（周六14:00)	2020-02-22	2020-06-20	36	2020-02-15	49	8	1
7	3163279	初一语文春季系统班（周六14:00)	2020-02-22	2020-06-20	36	2020-02-15	46	8	1
7	3163281	初一语文春季系统班（周六18:00)	2020-02-22	2020-06-20	36	2020-02-15	49	8	1
7	3163283	初一语文春季系统班（周六18:00)	2020-02-22	2020-06-20	36	2020-02-15	52	8	1
7	3163285	初一语文春季系统班（周日09:00)	2020-02-23	2020-06-21	36	2020-02-15	53	8	1
7	3163287	初一语文春季系统班（周日09:00)	2020-02-23	2020-06-21	36	2020-02-15	52	8	1
7	3163289	初一语文春季系统班（周日18:00)	2020-02-23	2020-06-21	36	2020-02-15	53	8	1
7	3511447	初一数学题型技巧特训班	2020-03-07	2020-03-10	7	2020-02-15	56	9	1
7	3168713	【目标A++班】初一数学春季系统班（周日14:00)	2020-02-23	2020-06-21	36	2020-02-15	57	9	1
7	3168709	【目标A++班】初一数学春季系统班（周六18:00)	2020-02-22	2020-06-20	36	2020-02-15	57	9	1
7	3168543	【目标A+班】初一数学春季系统班（人教版·周五18:00)	2020-02-21	2020-06-19	36	2020-02-15	59	9	1
7	3168545	【目标A+班】初一数学春季系统班（人教版·周五18:00)	2020-02-21	2020-06-19	36	2020-02-15	56	9	1
7	3168547	【目标A+班】初一数学春季系统班（人教版·周五18:00)	2020-02-21	2020-06-19	36	2020-02-15	61	9	1
7	3168549	【目标A+班】初一数学春季系统班（人教版·周五18:00)	2020-02-21	2020-06-19	36	2020-02-15	62	9	1
7	3168551	【目标A+班】初一数学春季系统班（人教版·周六09:00)	2020-02-22	2020-06-20	36	2020-02-15	61	9	1
7	3168553	【目标A+班】初一数学春季系统班（人教版·周六09:00)	2020-02-22	2020-06-20	36	2020-02-15	64	9	1
7	3168555	【目标A+班】初一数学春季系统班（人教版·周六09:00)	2020-02-22	2020-06-20	36	2020-02-15	65	9	1
7	3168557	【目标A+班】初一数学春季系统班（人教版·周六14:00)	2020-02-22	2020-06-20	36	2020-02-15	66	9	1
7	3168559	【目标A+班】初一数学春季系统班（人教版·周六14:00)	2020-02-22	2020-06-20	36	2020-02-15	56	9	1
7	3168561	【目标A+班】初一数学春季系统班（人教版·周六14:00)	2020-02-22	2020-06-20	36	2020-02-15	68	9	1
7	3168563	【目标A+班】初一数学春季系统班（人教版·周六14:00)	2020-02-22	2020-06-20	36	2020-02-15	69	9	1
7	3168565	【目标A+班】初一数学春季系统班（人教版·周六18:00)	2020-02-22	2020-06-20	36	2020-02-15	61	9	1
7	3168567	【目标A+班】初一数学春季系统班（人教版·周六18:00)	2020-02-22	2020-06-20	36	2020-02-15	64	9	1
7	3168569	【目标A+班】初一数学春季系统班（人教版·周六18:00)	2020-02-22	2020-06-20	36	2020-02-15	72	9	1
7	3168571	【目标A+班】初一数学春季系统班（人教版·周六18:00)	2020-02-22	2020-06-20	36	2020-02-15	59	9	1

b) sale info table:



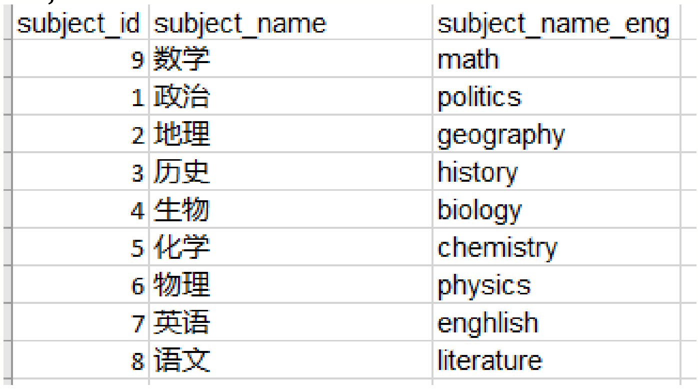
course_id	price	original_price	quantity	sold_count	pre_sale_time	start_sale_time	end_sale_time	left_count	sold_out
3511191	9	299	80	7			2020-03-01	73	0
3163273	1200	0	30	26	2019-10-26	2019-10-26	2020-04-07	4	0
3163275	1200	0	30	21	2019-10-26	2019-10-26	2020-04-07	9	0
3163277	1200	0	30	26	2019-10-26	2019-10-26	2020-04-07	4	0
3163279	1200	0	30	13	2019-10-26	2019-10-26	2020-04-07	17	0
3163281	1200	0	30	22	2019-10-26	2019-10-26	2020-04-07	8	0
3163283	1200	0	30	10	2019-10-26	2019-10-26	2020-04-07	20	0
3163285	1200	0	30	22	2019-10-26	2019-10-26	2020-04-07	8	0
3163287	1200	0	30	29	2019-10-26	2019-10-26	2020-04-07	1	0
3163289	1200	0	30	8	2019-10-26	2019-10-26	2020-04-07	22	0
3511447	9	299	80	12			2020-03-01	68	0
3168713	1200	0	30	24	2019-10-26	2019-10-26	2020-04-07	6	0
3168709	1200	0	30	23	2019-10-26	2019-10-26	2020-04-07	0	1
3168543	1200	0	30	16	2019-10-26	2019-10-26	2020-04-07	14	0
3168545	1200	0	30	20	2019-10-26	2019-10-26	2020-04-07	10	0
3168547	1200	0	30	21	2019-10-26	2019-10-26	2020-04-07	9	0
3168549	1200	0	30	16	2019-10-26	2019-10-26	2020-04-07	14	0
3168551	1200	0	30	16	2019-10-26	2019-10-26	2020-04-07	14	0
3168553	1200	0	30	24	2019-10-26	2019-10-26	2020-04-07	6	0
3168555	1200	0	30	23	2019-10-26	2019-10-26	2020-04-07	7	0

c) teacher table



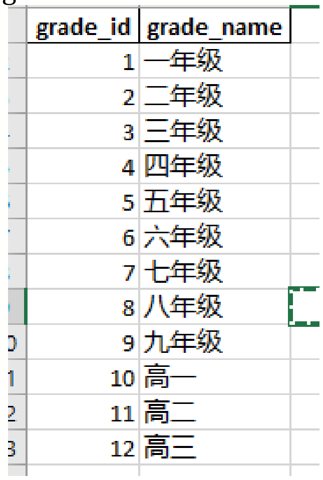
teacher_id	teacher_name
0	刘薇
1	程磊
2	严攀
4	斯琴
5	满晓桐
6	苗锟
7	周梦麟
8	梁冰
9	平赫
11	铁健栩
12	李林
15	刘和妍

d) subject table



subject_id	subject_name	subject_name_eng
9	数学	math
1	政治	politics
2	地理	geography
3	历史	history
4	生物	biology
5	化学	chemistry
6	物理	physics
7	英语	enghlish
8	语文	literature

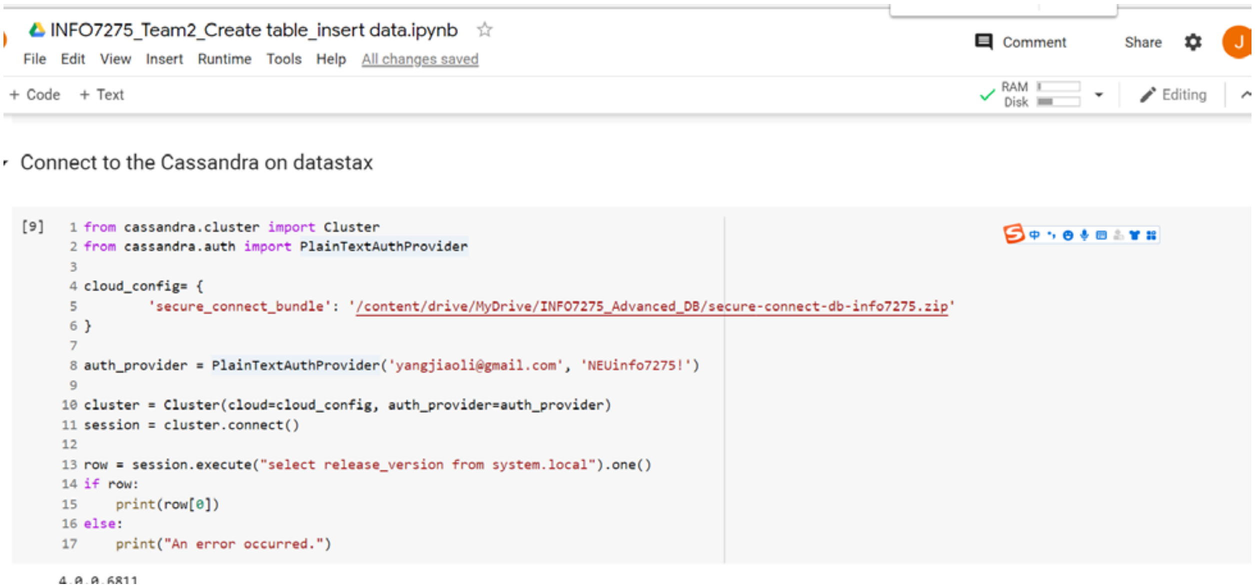
e) grade table



grade_id	grade_name
1	一年级
2	二年级
3	三年级
4	四年级
5	五年级
6	六年级
7	七年级
8	八年级
9	九年级
10	高一
11	高二
12	高三

5) Create table and Insert data

1. Connect to the Cassandra on datastax by using cassandra-driver

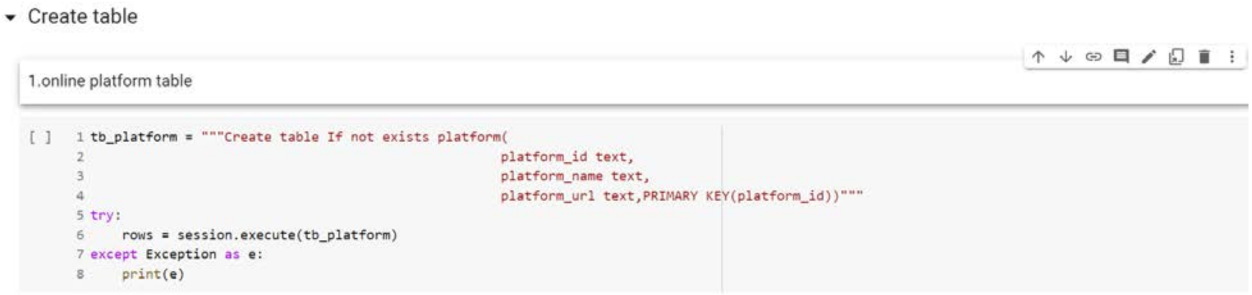

INFO7275_Team2_Create table_insert data.ipynb
File Edit View Insert Runtime Tools Help All changes saved
Comment
Share
J
+ Code + Text
RAM
Disk
Editing

Connect to the Cassandra on datastax

```
[9]
from cassandra.cluster import Cluster
from cassandra.auth import PlainTextAuthProvider

cloud_config= {
        'secure_connect_bundle': '/content/drive/MyDrive/INFO7275_Advanced_DB/secure-connect-db-info7275.zip'
}

auth_provider = PlainTextAuthProvider('yangjiaoli@gmail.com', 'NEUinfo7275!')

cluster = Cluster(cloud=cloud_config, auth_provider=auth_provider)
session = cluster.connect()

row = session.execute("select release_version from system.local").one()
if row:
    print(row[0])
else:
    print("An error occurred.")
```

```
4.0.0.6811
```

2.Create table

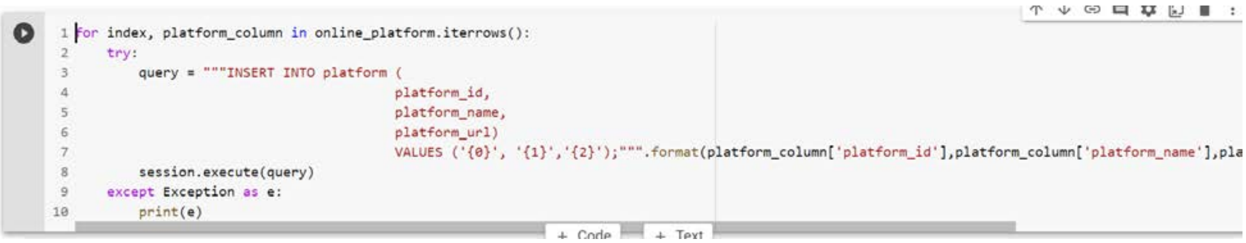


Create table

1.online platform table

```
[ ]
tb_platform = """Create table If not exists platform(
                                                    platform_id text,
                                                    platform_name text,
                                                    platform_url text,PRIMARY KEY(platform_id))"""
try:
    rows = session.execute(tb_platform)
except Exception as e:
    print(e)
```

3.Insert data

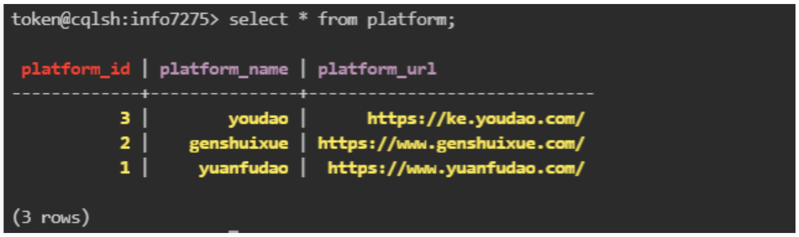


```
for index, platform_column in online_platform.iterrows():
    try:
        query = """INSERT INTO platform (
                                        platform_id,
                                        platform_name,
                                        platform_url)
                                        VALUES ('{0}', '{1}','{2}');""".format(platform_column['platform_id'],platform_column['platform_name'],pla
        session.execute(query)
    except Exception as e:
        print(e)
```

5) Check data in datastax

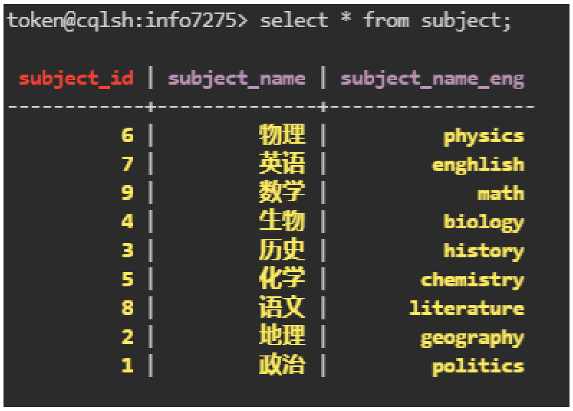
CQL console:

Platform table:



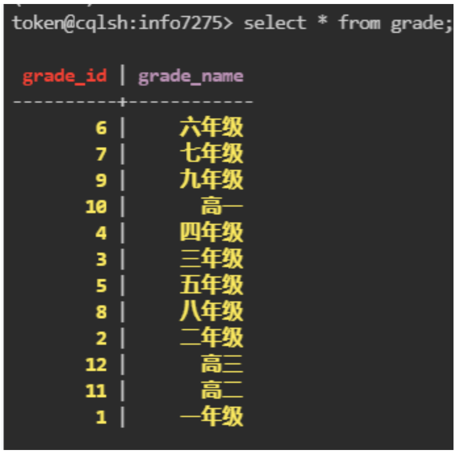
```
token@cqlsh:info7275> select * from platform;

 platform_id | platform_name | platform_url
-------------+---------------+----------------------------
           3 |        youdao |       https://ke.youdao.com/
           2 |    genshuixue | https://www.genshuixue.com/
           1 |     yuanfudao |  https://www.yuanfudao.com/

(3 rows)
```

Subject table:



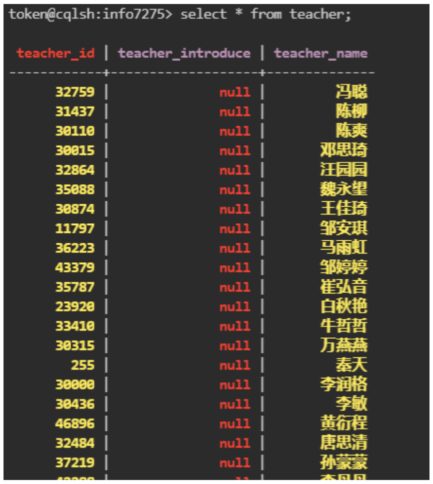
```
token@cqlsh:info7275> select * from subject;

 subject_id | subject_name | subject_name_eng
------------+--------------+------------------
          6 |         物理 |          physics
          7 |         英语 |         enghlish
          9 |         数学 |             math
          4 |         生物 |          biology
          3 |         历史 |          history
          5 |         化学 |        chemistry
          8 |         语文 |       literature
          2 |         地理 |        geography
          1 |         政治 |         politics
```

Grade table:



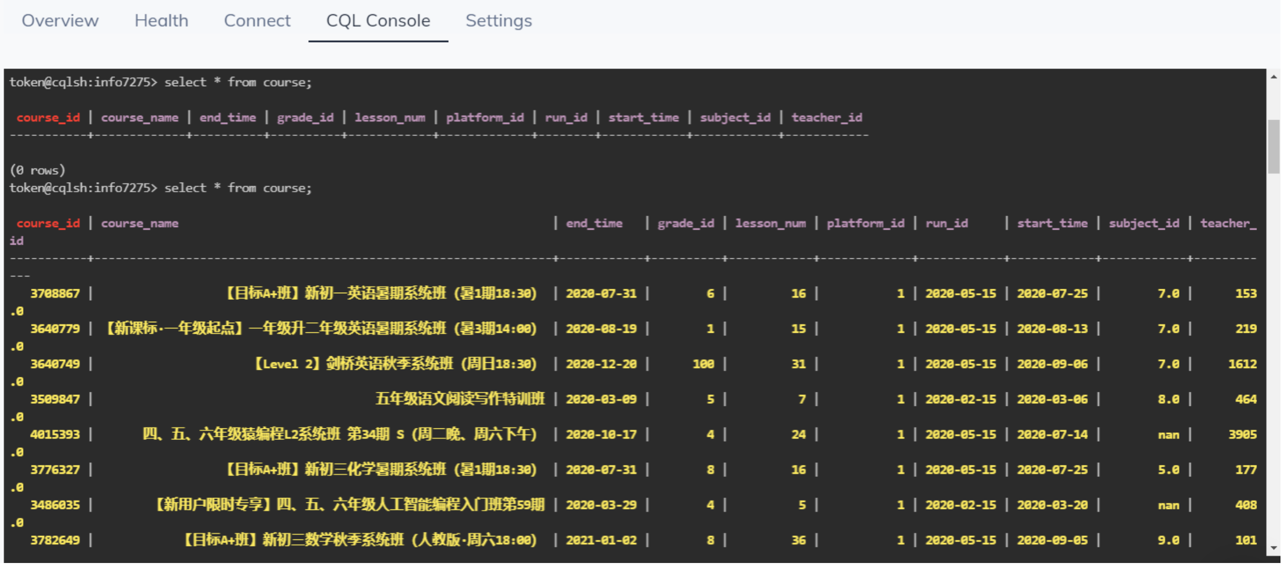
```
token@cqlsh:info7275> select * from grade;

 grade_id | grade_name
----------+------------
        6 |     六年级
        7 |     七年级
        9 |     九年级
       10 |       高一
        4 |     四年级
        3 |     三年级
        5 |     五年级
        8 |     八年级
        2 |     二年级
       12 |       高三
       11 |       高二
        1 |     一年级
```

Teacher table:



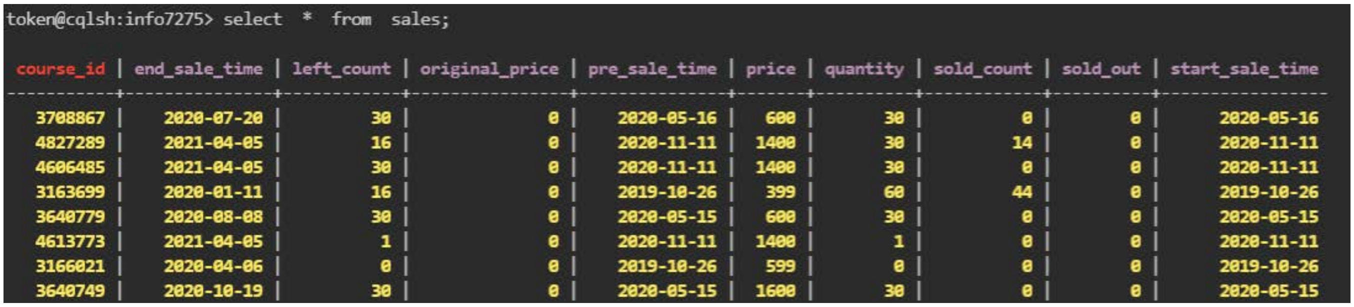
```
token@cqlsh:info7275> select * from teacher;

 teacher_id | teacher_introduce | teacher_name
------------+-------------------+--------------
      32759 |              null |         冯聪
      31437 |              null |         陈柳
      30110 |              null |         陈爽
      30015 |              null |       邓思琦
      32864 |              null |       汪园园
      35088 |              null |       魏永望
      30874 |              null |       王佳琦
      11797 |              null |       邹安琪
      36223 |              null |       马雨虹
      43379 |              null |       邹婷婷
      35787 |              null |       崔弘音
      23920 |              null |       白秋艳
      33410 |              null |       牛哲哲
      30315 |              null |       万燕燕
        255 |              null |         秦天
      30000 |              null |       李润格
      30436 |              null |         李敏
      46896 |              null |       黄衍程
      32484 |              null |       唐思清
      37219 |              null |       孙蒙蒙
```

Course table:



```
token@cqlsh:info7275> select * from course;

 course_id | course_name | end_time | grade_id | lesson_num | platform_id | run_id | start_time | subject_id | teacher_id
-----------+-------------+----------+----------+------------+-------------+--------+------------+------------+------------

(0 rows)
token@cqlsh:info7275> select * from course;

 course_id | course_name                                        | end_time   | grade_id | lesson_num | platform_id | run_id     | start_time | subject_id | teacher_
id
-----------+----------------------------------------------------+------------+----------+------------+-------------+------------+------------+------------+---------
---
   3708867 |            【目标A+班】新初一英语暑期系统班（暑1期18:30） | 2020-07-31 |        6 |         16 |           1 | 2020-05-15 | 2020-07-25 |        7.0 |      153
.0
   3640779 | 【新课标·一年级起点】一年级升二年级英语暑期系统班（暑3期14:00） | 2020-08-19 |        1 |         15 |           1 | 2020-05-15 | 2020-08-13 |        7.0 |      219
.0
   3640749 |               【Level 2】剑桥英语秋季系统班（周日18:30） | 2020-12-20 |      100 |         31 |           1 | 2020-05-15 | 2020-09-06 |        7.0 |     1612
.0
   3509847 |                           五年级语文阅读写作特训班 | 2020-03-09 |        5 |          7 |           1 | 2020-02-15 | 2020-03-06 |        8.0 |      464
.0
   4015393 |    四、五、六年级猿编程L2系统班 第34期 s（周二晚、周六下午） | 2020-10-17 |        4 |         24 |           1 | 2020-05-15 | 2020-07-14 |        nan |     3905
.0
   3776327 |            【目标A+班】新初三化学暑期系统班（暑1期18:30） | 2020-07-31 |        8 |         16 |           1 | 2020-05-15 | 2020-07-25 |        5.0 |      177
.0
   3486035 |   【新用户限时专享】四、五、六年级人工智能编程入门班第59期 | 2020-03-29 |        4 |          5 |           1 | 2020-02-15 | 2020-03-20 |        nan |      408
.0
   3782649 |      【目标A+班】新初三数学秋季系统班（人教版·周六18:00） | 2021-01-02 |        8 |         36 |           1 | 2020-05-15 | 2020-09-05 |        9.0 |      101
```

Sales table：



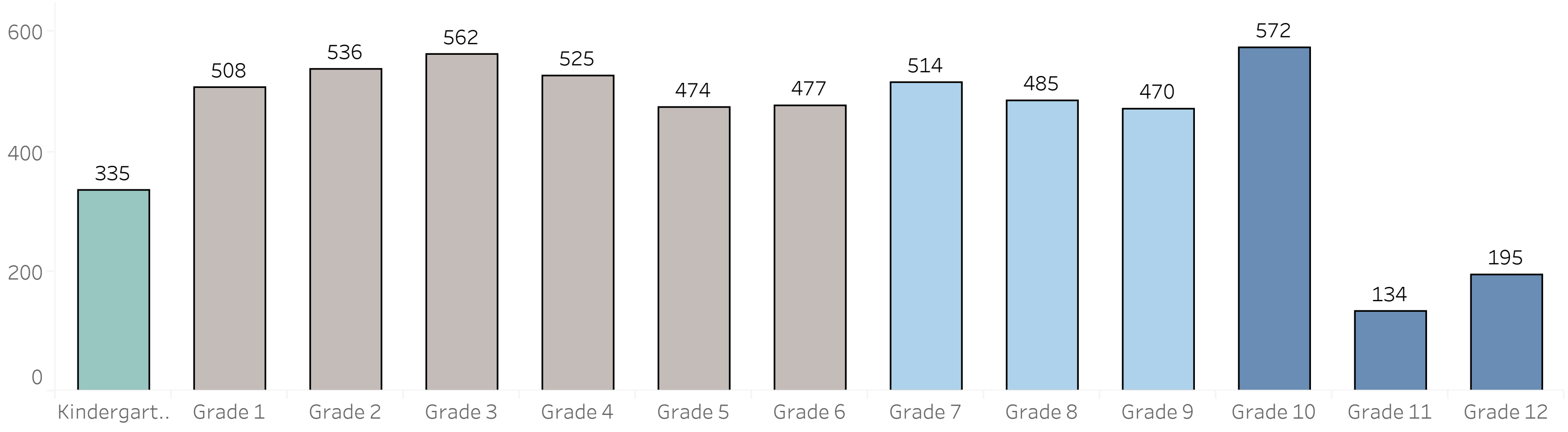
```
token@cqlsh:info7275> select  *  from  sales;

 course_id | end_sale_time | left_count | original_price | pre_sale_time | price | quantity | sold_count | sold_out | start_sale_time
-----------+---------------+------------+----------------+---------------+-------+----------+------------+----------+-----------------
   3708867 |    2020-07-20 |         30 |              0 |    2020-05-16 |   600 |       30 |          0 |        0 |      2020-05-16
   4827289 |    2021-04-05 |         16 |              0 |    2020-11-11 |  1400 |       30 |         14 |        0 |      2020-11-11
   4606485 |    2021-04-05 |         30 |              0 |    2020-11-11 |  1400 |       30 |          0 |        0 |      2020-11-11
   3163699 |    2020-01-11 |         16 |              0 |    2019-10-26 |   399 |       60 |         44 |        0 |      2019-10-26
   3640779 |    2020-08-08 |         30 |              0 |    2020-05-15 |   600 |       30 |          0 |        0 |      2020-05-15
   4613773 |    2021-04-05 |          1 |              0 |    2020-11-11 |  1400 |        1 |          0 |        0 |      2020-11-11
   3166021 |    2020-04-06 |          0 |              0 |    2019-10-26 |   599 |        0 |          0 |        0 |      2019-10-26
   3640749 |    2020-10-19 |         30 |              0 |    2020-05-15 |  1600 |       30 |          0 |        0 |      2020-05-15
```

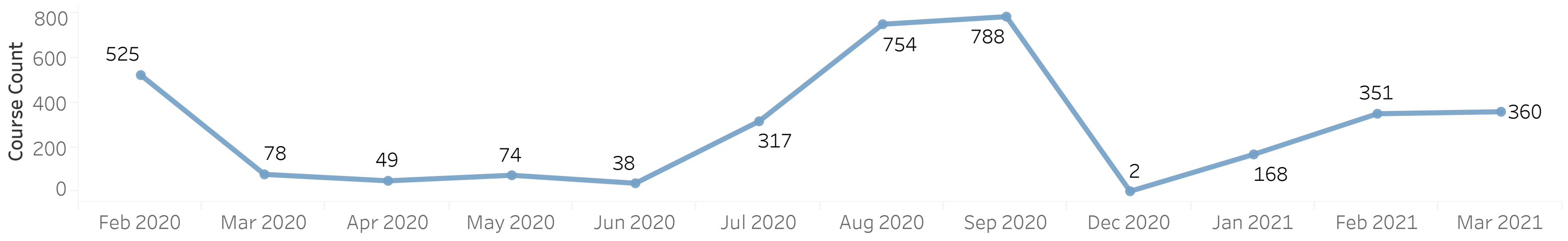
Part 4:

Data Visualization

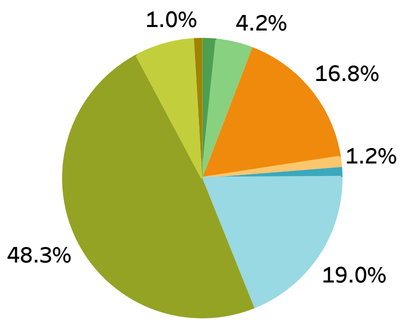
China’s K12 Online Education Data Management System

Course Count by Grade

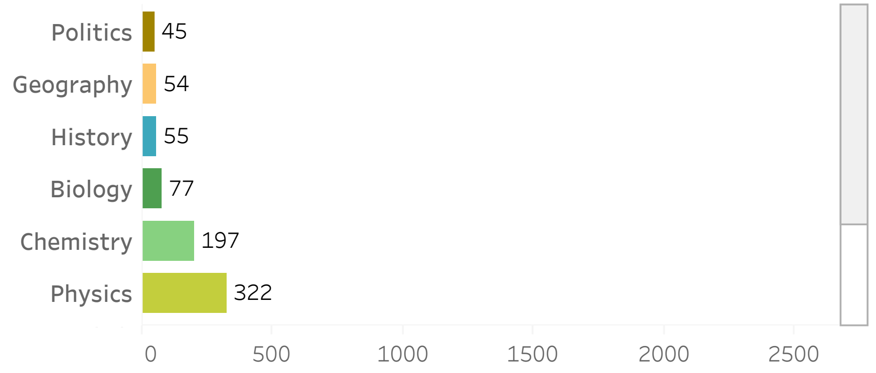




Course Count Percent By Subject

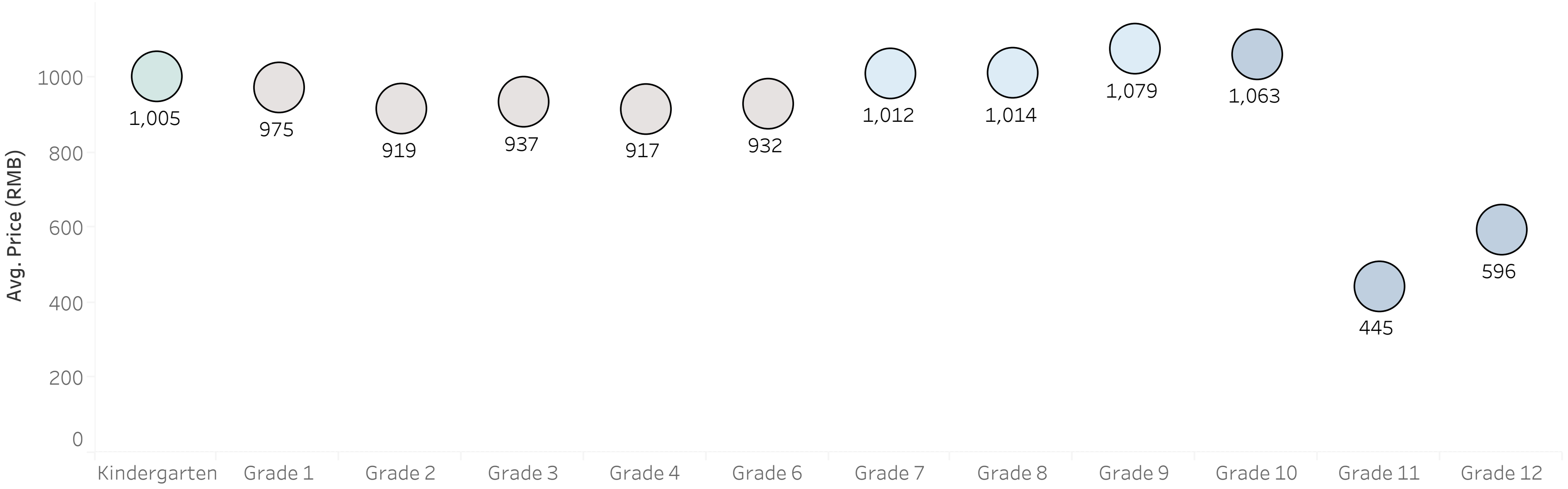


Course Count By Subject



China’s K12 Online Education Data Management System

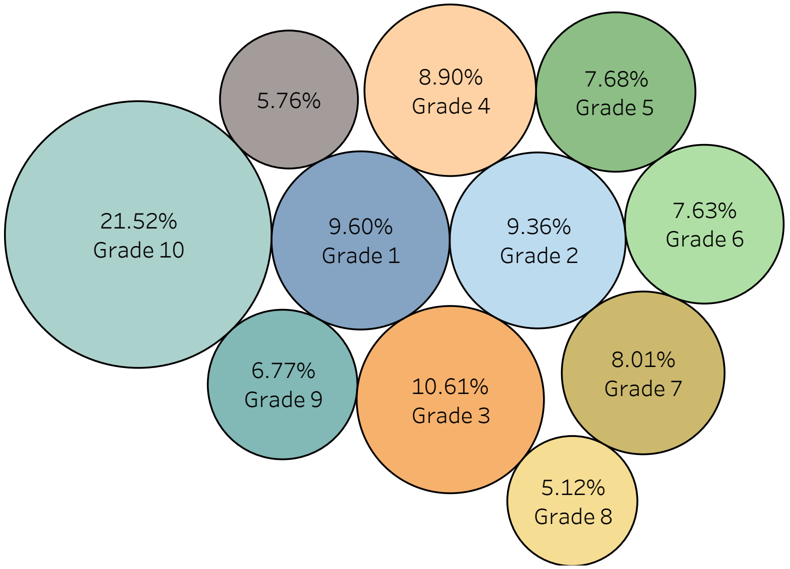
Course Average Price by Grade



Course Sale Infomation

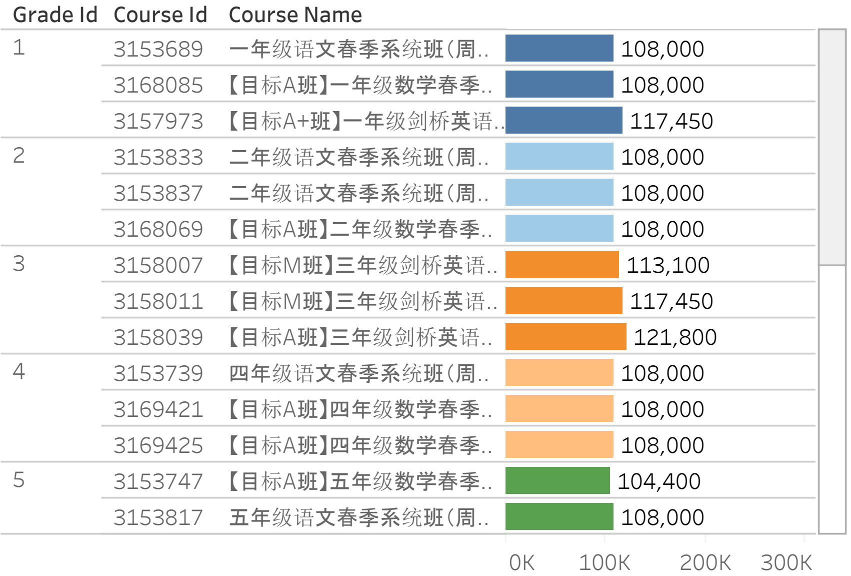
	Avg Price	Course Qty	Sold Qty	Pct of Sold
Biology	953	2,830	1,493	23.6%
Chemistry	1,003	6,850	4,135	26.4%
English	983	30,268	13,669	24.4%
Geography	999	2,230	743	14.8%
History	977	2,290	1,102	20.4%
Literature	936	32,529	15,694	25.1%
Math	980	77,836	52,104	40.2%
Physics	1,004	11,870	5,463	21.9%
Politics	971	1,810	455	10.1%

Course Revenue By Grade



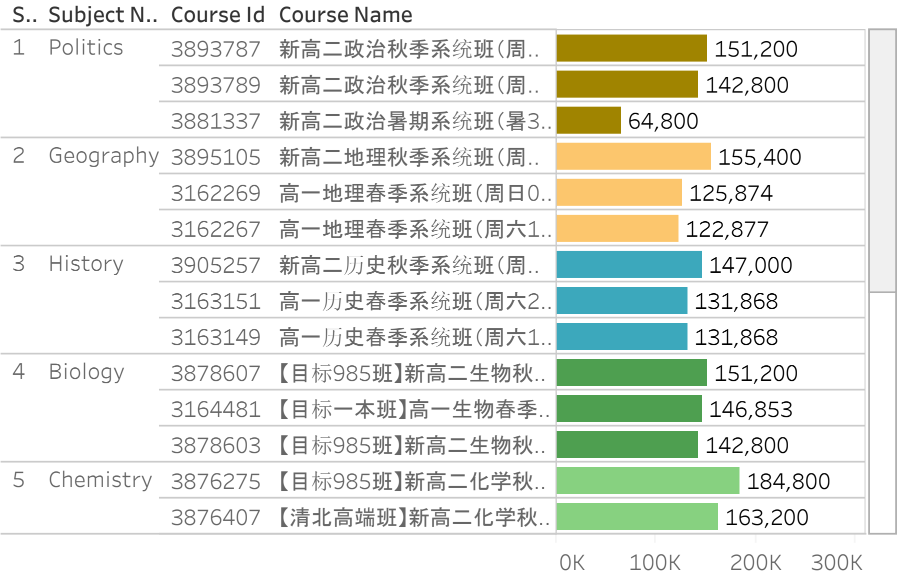
China’s K12 Online Education Data Management System

Top 3 course each grade ranked by revenue



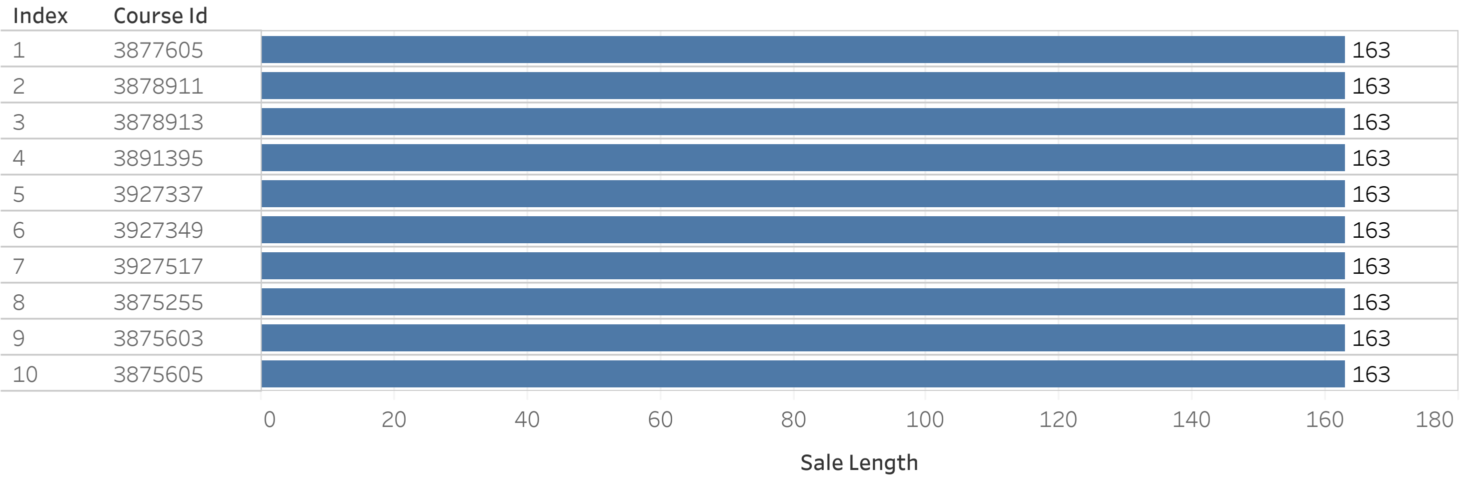
Grade Id	Course Id	Course Name	
1	3153689	一年级语文春季系统班(周..	108,000
	3168085	【目标A班】一年级数学春季..	108,000
	3157973	【目标A+班】一年级剑桥英语..	117,450
2	3153833	二年级语文春季系统班(周..	108,000
	3153837	二年级语文春季系统班(周..	108,000
	3168069	【目标A班】二年级数学春季..	108,000
3	3158007	【目标M班】三年级剑桥英语..	113,100
	3158011	【目标M班】三年级剑桥英语..	117,450
	3158039	【目标A班】三年级剑桥英语..	121,800
4	3153739	四年级语文春季系统班(周..	108,000
	3169421	【目标A班】四年级数学春季..	108,000
	3169425	【目标A班】四年级数学春季..	108,000
5	3153747	【目标A班】五年级数学春季..	104,400
	3153817	五年级语文春季系统班(周..	108,000

Top 3 course each subject ranked by revenue



S..	Subject N..	Course Id	Course Name	
1	Politics	3893787	新高二政治秋季系统班(周..	151,200
		3893789	新高二政治秋季系统班(周..	142,800
		3881337	新高二政治暑期系统班(暑3..	64,800
2	Geography	3895105	新高二地理秋季系统班(周..	155,400
		3162269	高一地理春季系统班(周日0..	125,874
		3162267	高一地理春季系统班(周六1..	122,877
3	History	3905257	新高二历史秋季系统班(周..	147,000
		3163151	高一历史春季系统班(周六2..	131,868
		3163149	高一历史春季系统班(周六1..	131,868
4	Biology	3878607	【目标985班】新高二生物秋..	151,200
		3164481	【目标一本班】高一生物春季..	146,853
		3878603	【目标985班】新高二生物秋..	142,800
5	Chemistry	3876275	【目标985班】新高二化学秋..	184,800
		3876407	【清北高端班】新高二化学秋..	163,200

Top10 course by longest sale time(days)



Index	Course Id	
1	3877605	163
2	3878911	163
3	3878913	163
4	3891395	163
5	3927337	163
6	3927349	163
7	3927517	163
8	3875255	163
9	3875603	163
10	3875605	163