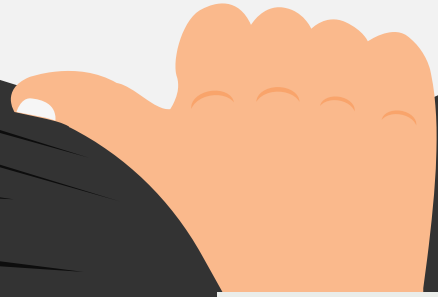


- '서울시 커피숍 정보와

서울시 문화공간 현황 DATA 상권 분석' -



20142770

서울시 커피숍 현황

번호	사업장명	소재지전체주소	도로명전체주소	인허가일자	영업상태명	폐업일자	휴업시작일자
1	곰미커피	서울특별시 성동구 성수동1가 6...	서울특별시 성동구 서울숲4길 2...	20130605	운영중		
2	학다방	서울특별시 성동구 행당동 196...	서울특별시 성동구 왕십리로 29...	19760205	운영중		
3	남탕매점	서울특별시 성동구 성수동2가 2...	서울특별시 성동구 성수이로 11...	20030315	운영중		
4	솔	서울특별시 성동구 용답동 234...	서울특별시 성동구 자등차시장1...	19810723	운영중		
5	아람다방	서울특별시 성동구 성수동1가 1...	서울특별시 성동구 연무장길 6 ...	19871022	운영중		
6	우회	서울특별시 성동구 금호동3가 4...	서울특별시 성동구 독서당로 29...	19760205	운영중		
7	삼거리	서울특별시 성동구 도선동 363...	서울특별시 성동구 왕십리로 36...	19760422	운영중		
8	대지다방	서울특별시 성동구 용답동 234...	서울특별시 성동구 자등차시장1...	19911116	운영중		
9	수정	서울특별시 성동구 용답동 234...		19850716	운영중		

번호/사업장/소재지 주소/도로 명 주소/ 인허가 일자/ 영업 상태의 순으로 저장되어 있음

→ 6개 FIELD

서울시 문화공간 현황

문화공간코드	장르분류코드	장르분류명	문화공간명	대표이미지	주소	전화번호	팩스
100319	3	박물관/기념관	한상수자수박물관	http://culture.seoul.go.kr/d...	서울 종로구 가회동 11-32)	02-744-1545	
100517	1	공연장	서울에니시네마	http://culture.seoul.go.kr/d...	서울 중구 예장동 8-145)	02-3455-8318	02-34...
100464	8	도서관	LG상남도서관	http://culture.seoul.go.kr/d...	서울 종로구 원서동 창덕궁 1...	02-708-3700	02-70...
100082	8	도서관	강서도서관	http://culture.seoul.go.kr/d...	서울 강서구 등촌동 520-6)	02-3219-7000	02-26...
100873	1	공연장	명동난타극장	http://culture.seoul.go.kr/d...	서울 중구 명동길 26	02-739-8288	
100457	1	공연장	연강홀	http://culture.seoul.go.kr/d...	서울 종로구 연지동 270)	02-708-5001	02-70...
100813	8	도서관	금천구립정보도서관	http://culture.seoul.go.kr/d...	서울 금천구	02-863-9544~6	02-86...
100877	6	문화예술화관	북촌 서울무형문화재 교육전...	http://culture.seoul.go.kr/d...	서울 종로구 재동 53-1)	02-747-0303	02-74...
100311	7	문화원	독일문화원	http://culture.seoul.go.kr/d...	서울 용산구 후암동 339-1)	02-2021-2800	

문화공간 코드/ 장르 분류 코드/ 장르 분류 명/ 문화 공간 명/ 대표 이미지/ 주소/ 전화/ 팩스/ 홈페이지/ 관람시간 / 관람료/ 휴관 일 / 개관일자/ 객석 수/ X좌표 /Y좌표 / 기타 사항/ 무료 유료 구분

→ 18개 FIELD

MR 작업 추가



SPARK MLIB 의미 있는 활용

Enjoy your stylish business
and campus life with BIZCAM

→ MR 작업을 통해 구 별 커피숍 정보와 문화 공간 개수를 추출

→ (구 별 커피숍 개수, 구 별 문화 공간 개수)를 입력으로 SPARK의 MLIB 中 PEARSON CORRELATION

SPARK (MLIB)

SCALA (MLIB)

GOOGLE CLOUD PLATFORM 컴퓨팅 환경

HADOOP (MAP-REDUCE)

JAVA (MR)

MR 작업 과정

```
import java.io.IOException;
import org.apache.hadoop.conf.Configuration;
import org.apache.hadoop.fs.Path;
import org.apache.hadoop.io.IntWritable;
import org.apache.hadoop.io.LongWritable;
import org.apache.hadoop.io.Text;
import org.apache.hadoop.mapreduce.Job;
import org.apache.hadoop.mapreduce.Mapper;
import org.apache.hadoop.mapreduce.Reducer;
import org.apache.hadoop.mapreduce.lib.input.FileInputFormat;
import org.apache.hadoop.mapreduce.lib.input.TextInputFormat;
import org.apache.hadoop.mapreduce.lib.output.FileOutputFormat;
import org.apache.hadoop.mapreduce.lib.output.TextOutputFormat;
import java.util.Arrays;
import java.util.ArrayList;

public class infra_per_ku {

    public static void main(String[] args) throws Exception{

        Configuration conf=new Configuration();

        Job job = Job.getInstance(conf,"infra_per_ku");

        job.setJarByClass(infra_per_ku.class);

        // job.setNumReduceTasks(1);//To SUM "FILM-NOIR" GENRE TOGETHER AT ONE-REDUCE

        job.setOutputKeyClass(Text.class); //KEY-VALUE (MOVIE_TITLE - COUNT)
        job.setOutputValueClass(IntWritable.class);

        job.setMapperClass(infra_per_ku_MAP.class);
        job.setReducerClass(infra_per_ku_REDUCE.class);

        job.setInputFormatClass(TextInputFormat.class);
        job.setOutputFormatClass(TextOutputFormat.class);

        FileInputFormat.addInputPath(job,new Path(args[0]));
        FileOutputFormat.setOutputPath(job,new Path(args[1]));

        job.waitForCompletion(true);
    }
}
```

- 문화 시설 COUNT MR 코드. . COFFE SHOP 정보도 위와 동일
- 수업 시간에 했던 방식을 응용했지만 ERROR 발생했었다

```
public static class infra_per_ku_MAP extends Mapper<LongWritable,Text,Text,IntWritable> {

    private final static IntWritable ONE=new IntWritable(1);
    private Text word=new Text();

    public void map(LongWritable key,Text value,Context context)throws IOException,InterruptedException{
        String line=value.toString(); //READ LINE-BY-LINE
        String [] fields=line.split(",");//DECODE -CSV [" ",""," .....]
        if(fields.length>5){
            String address=fields[5];
            if (address!=null){
                String [] add=address.split(" ");
                if(add.length>2){
                    String for_key=add[1];
                    word.set(for_key);
                    context.write(word,ONE);
                }
            }
        }
    }
}

public static class infra_per_ku_REDUCE extends Reducer<Text,IntWritable,Text,IntWritable> {

    /*
     * REDUCE RETURN UNIQUE TITLE-ONE PER BLOCK. SO WE SHOULD GET UNIQUE IN REDUCE
     */

    public void reduce(Text key,Iterable<IntWritable> values,Context context) throws IOException,InterruptedException{

        int sum=0;

        for (IntWritable val:values) {

            sum+=val.get();

        }

        context.write(key,new IntWritable(sum));

    }
}
```

MR 작업 과정

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
2	100966	1	공연장	대학로예술극장	서울 종로	02-3668-0007									37.58197	127.0033		
3	100371	1	공연장	대학로문화	http://cult	서울 종로	02-785-2305			1F 원 (10:30 ~ 23:00	공휴일				37.5823	127.0037		
4	100799	1	공연장	신월문화	http://cult	서울 양천	02-2605-402-2606-4	http://www.ycs.or.kr							37.51985	126.835		
5	100913	1	공연장	(재)마포문화재단	서울 마포	02-3274-8633		http://www.mapoartcenter.or.k	매월 마지막주 월요일						37.54989	126.9453		
6	101145	1	공연장	그랜드 컨퍼런스룸	서울특별시	02-6000-0114									37.51186	127.0602		
7	100955	1	공연장	세종대 컨벤션센터	서울 광진	2.34E+08									37.55034	127.0732		
8	100699	1	공연장	청계천박	http://cult	서울 성동	02-2286-302-2286-3	http://www.cgcm.go.kr							37.57123	127.0349		
9	100161	1	공연장	동방예술	http://cult	서울 마포	02-712-38707-1132								37.55451	126.9385		
10	100931	1	공연장	JCC 아트	http://cult	서울 종로	02-2138-7372	www.jeijcc.org							37.58747	127.0023		
11	101203	1	공연장	의정부예	http://cult	경기도 의	031-826-5841	http://www.uac.or.kr/							37.73339	127.0341		
12	100012	1	공연장	CMG섬유	http://cult	서울 강남	02-528-4098~9	http://www.culturem.com/event/index.php?mode=info_photo_							37.50769	127.0604		
13	100132	1	공연장	롯데시네	http://cult	서울	02-1644-8855	http://www.lottecinema.co.kr/LHS/LHFS/Global/Main/Index.asp							37.65517	127.061		
14	101214	1	공연장	플랫폼 창	http://cult	culture.seoul	02-993-0561	www.platform61.kr/							37.65437	127.0491		

→ 주소 FILED를 공백 기준 SPLIT('WWS')해서 <'구',ONE> 를 KEY-VALUE로 MAP의 결과를 쓰려고 함.

→ 수업 시간에 진행했던 DATA와 달리 KEY로 쓰려는 값이 NULL 인 경우도 있고 형식이 다른 경우도 있었다.

MR 작업 과정

```
public static class infra_per_ku_MAP extends Mapper<LongWritable,Text,Text,IntWritable> {

    private final static IntWritable ONE=new IntWritable(1);
    private Text word=new Text();

    public void map(LongWritable key,Text value,Context context)throws IOException,InterruptedException{
        String line=value.toString();//READ LINE-BY-LINE
        String [] fields=line.split(",");//DECODE -CSV [" "," "," ".....]
        if(fields.length>5){
            String address=fields[5];
            if (address!=null){
                String [] add=address.split("\\s");
                if(add.length>2){
                    String for_key=add[1];
                    word.set(for_key);
                    context.write(word,ONE);
                }
            }
        }
    }
}

public static class infra_per_ku_REDUCE extends Reducer<Text,IntWritable,Text,IntWritable> {

    /*
     * REDUCE RETURN UNIQUE TITLE-ONE PER BLOCK. SO WE SHOULD GET UNIQUE IN REDUCE
     */

    public void reduce(Text key,Iterable<IntWritable> values,Context context) throws IOException,InterruptedException{

        int sum=0;

        for (IntWritable val:values) {

            sum+=val.get();

        }
        context.write(key,new IntWritable(sum));
    }
}
```

데이터가 일관되지 않는 경우 처리

→ 길이 혹은 공백을 기준으로 OPTION을 줌.

MR 작업 결과

```
choirak0805@kmbigdata-cluster-m:~/project/cafe$ hdfs dfs -cat /infra_result/* | sort -n -k2
18/12/07 09:21:00 INFO gcs.GoogleHadoopFileSystemBase: GHFS version: 1.6.10-hadoop2
522
의정부시 1
구현면로 1
남양주시 1
포천시 1
표천시 1
안산시 1
양천구 1
관악구 1
강서구 1
구연시 1
하남시 1
고양시 1
구리시 1
통각구 1
연주호 1
군포시 1
16 1
일원역339-3호(23번출입구 1
9 1
강북구 2
종로구 2
성남시 2
인양시 2
영등포구 3
도봉구 3
강동구 3
영북구 3
노원구 3
종아문구 4
금천구 4
마포구 5
종파구 5
성동구 5
서대문구 7
광진구 7
서초구 8
은평구 12
구로구 12
종로구 17
강남구 29
중구 32
영산구 66
choirak0805@kmbigdata-cluster-m:~/project/cafe$
```

→ 구 별 문화 공간 개수

```
choirak0805@kmbigdata-cluster-m:~/project/cafe$ hdfs dfs -cat /cafe_result/* | sort -n -k2
18/12/07 09:19:30 INFO gcs.GoogleHadoopFileSystemBase: GHFS version: 1.6.10-hadoop2
도봉구 3
양천구 5
강동구 7
광진구 10
은평구 11
관악구 12
성동구 12
통각구 12
종량구 14
강북구 15
송파구 18
노원구 19
영산구 21
성북구 21
구로구 25
강서구 29
마포구 32
종대문구 36
서초구 36
금천구 40
서대문구 41
영등포구 95
종로구 138
중구 270
강남구 635
```

→ 구 별 카페 개수

SPARK MLIB를 中 PEARSON CORRELATION

```
choirak0805@kmubiqdata-cluster-m: ~ - Chrome
https://ssh.cloud.google.com/projects/sound-decoder-221702/zones/asia-northeast1-c/instances/ki

Type in expressions to have them evaluated.
Type :help for more information.

scala> sc
res0: org.apache.spark.SparkContext = org.apache.spark.SparkContext@47f39279

scala> import org.apache.spark.mllib.linalg._
import org.apache.spark.mllib.linalg._

scala> import org.apache.spark.mllib.stat.Statistics
import org.apache.spark.mllib.stat.Statistics

scala> val sp=sc.parallelize(List(
  | Vectors.dense(3,3),
  | Vectors.dense(5,1),
  | Vectors.dense(7,3),
  | Vectors.dense(10,7),
  | Vectors.dense(11,12),
  | Vectors.dense(12,1),
  | Vectors.dense(12,5),
  | Vectors.dense(12,1),
  | Vectors.dense(14,2),
  | Vectors.dense(15,2),
  | Vectors.dense(18,5),
  | Vectors.dense(19,3),
  | Vectors.dense(21,66),
  | Vectors.dense(21,3),
  | Vectors.dense(25,12),
  | Vectors.dense(29,1),
  | Vectors.dense(32,5),
  | Vectors.dense(36,4),
  | Vectors.dense(36,8),
  | Vectors.dense(40,4),
  | Vectors.dense(41,7),
  | Vectors.dense(95,3),
  | Vectors.dense(138,17),
  | Vectors.dense(270,32),
  | Vectors.dense(635,29)
  | ))
sp: org.apache.spark.rdd.RDD[org.apache.spark.mllib.linalg.Vector] = ParallelCollectionRDD[0] at
parallelize at <console>:28

scala> val corr = Statistics.corr(sp)
corr: org.apache.spark.mllib.linalg.Matrix =
1.0          0.4050482853110014
0.4050482853110014  1.0
```

→ 서울시 문화공간과 카페의 상관 계수 분석을 위해 MR 작업에서 도출한 25개 구의 (구 별 커피숍 개수, 구 별 문화공간 개수) 를 입력으로 PEARSON CORRELATION 적용

- WHY PEARSON CORRELATION?
- 변수 간 관련도 → CORRELATION 분석법 사용
- 몇 가지 방식이 있는데 SPEARMAN CORREALATION 은 두 변수가 순위 형태일 때 사용
- COEFFICIEINT는 -1~ 1이고 그 값으로 상관 정도 해석

```
corr: org.apache.spark.mllib.linalg.Matrix =
1.0          0.4050482853110014
0.4050482853110014  1.0
```

→ 그 결과 상관 계수가 0.4050으로 : 강력한 양(+)의 상관. 변인 x 가 증가하면 변인 y 가 증가한다고 판단 가능하다는 결과 도출.

→ 여기서는 커피숍개수와 문화공간 개수가 양의 상관 관계를 가진다.