

다 른 조 는 몰 라 도 5 조 는 영 원 히

# Hadoop & Ecosystem

5조 김용선 문소희 박세영 박하얀 서영철 정아인

# CONTENTS

01

프로젝트  
목표

02

적용 기술

03

작업 과정

04

구현 내역

01

프로젝트 목표:

# 2019년 국가별 행복지수 데이터 수집 및 분석

# 02

## 적용기술(아키텍처)



# 03

## 작업과정 - 데이터 수집(kaggle)

≡ kaggle

🔍 Search

- 🏠 Home
- 🏆 Compete
- 📊 Data
- 📁 Notebooks
- 💬 Discuss
- 📦 Courses
- ⌵ More

Recently Viewed

- 🖼️ Netflix Movies and TV ...
- 🖼️ World Happiness Report

📁 Dataset



1890

### World Happiness Report

Happiness scored according to economic production, social support, etc.



Sustainable Development Solutions Network • updated 7 months ago (Version 2)

[Data](#) [Tasks \(1\)](#) [Kernels \(677\)](#) [Discussion \(10\)](#) [Activity](#) [Metadata](#)

Download (79 KB)

[New Notebook](#)

📊 Usability 8.5

📜 License CC0: Public Domain

🏷️ Tags arts and entertainment, statistics, reference, news, social sciences and 4 more

Description

#### Context

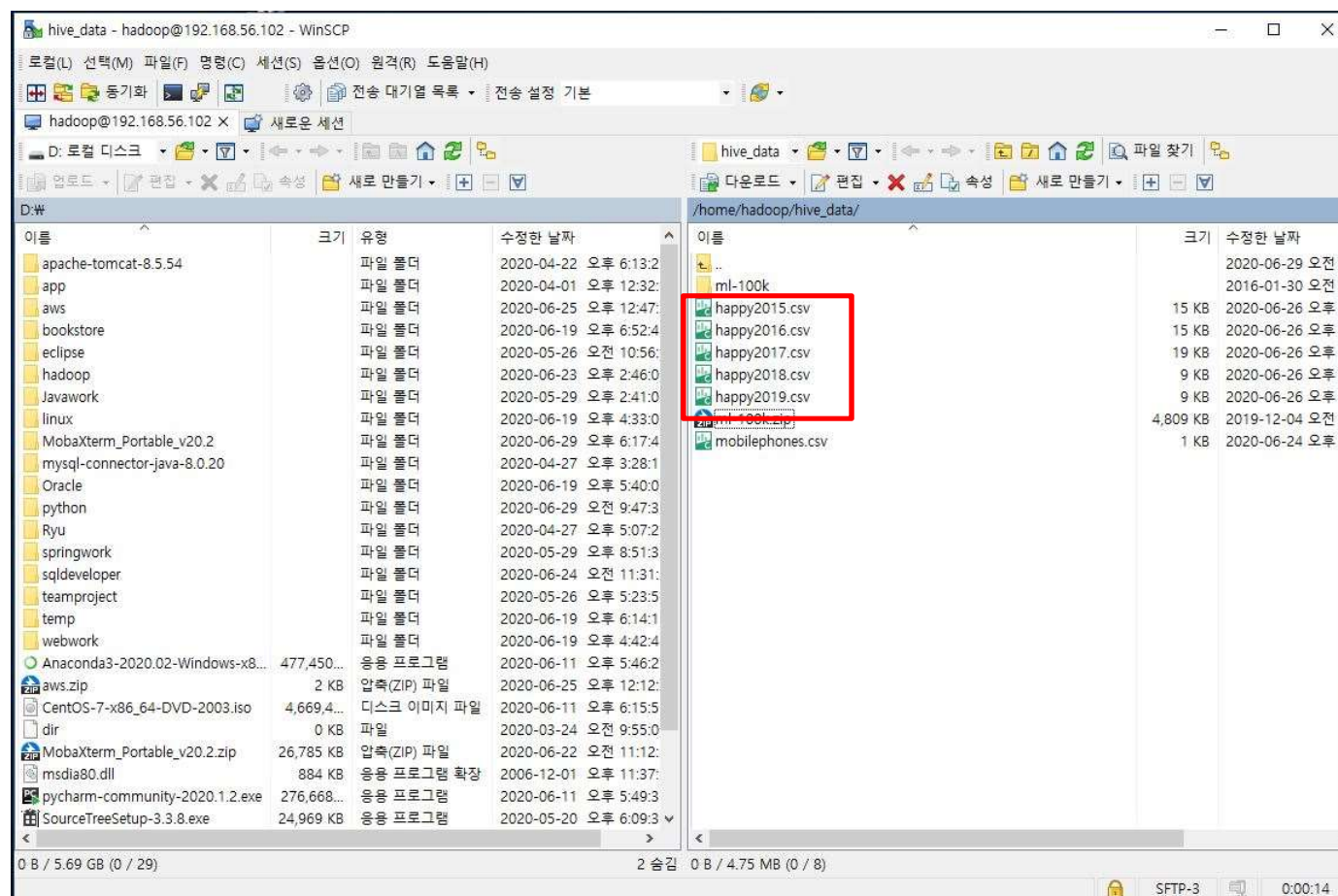
The World Happiness Report is a landmark survey of the state of global happiness. The first report was published in 2012, the second in 2013, the third in 2015, and the fourth in the 2016 Update. The World Happiness 2017, which ranks 155 countries by their happiness levels, was released at the United Nations at an event celebrating International Day of Happiness on March 20th. The report continues to gain global recognition as governments, organizations and civil society increasingly use happiness indicators to inform their policy-making decisions. Leading experts across fields – economics, psychology, survey analysis, national statistics, health, public policy and more – describe how measurements of well-being can be used effectively to assess the progress of nations. The reports review the state of happiness in the world today and show how the new science of happiness explains personal and national variations in happiness.

Content

# 03

## 작업과정 - 데이터 수집

Win SCP를 통해 Linux시스템으로 이동  
(경로: /home/Hadoop/hive\_data/)



# 03

## 작업과정 - 데이터 처리(Hive)

### 1. 데이터 처리를 위한 Database 생성

```
hive (default)> create database happiness;
```

### 2. 생성 확인

```
hive (default)> show databases;  
OK  
default  
happiness  
hivedemo  
kdatademo  
sample1  
Time taken: 5.854 seconds, Fetched: 5 row(s)
```

### 3. Database 사용

```
hive (default)> use happiness;
```

# 03

## 작업과정 - 데이터 처리(Hive)

### 1. 테이블 생성

```
hive (happiness)> create table happy2019(  
    > hrank Int,  
    > country string,  
    > hscore float,  
    > gdp float,  
    > family float,  
    > healthy float,  
    > freedom float,  
    > trust float,  
    > generosity float)  
    > ROW FORMAT DELIMITED FIELDS TERMINATED BY ','  
    > STORED AS TEXTFILE;
```

### 2. Hive에서 사용 가능하도록 테이블에 load

```
hive (happiness)> load data local inpath  
    > '/home/hadoop/hive_data/happy2019.csv'  
    > into table happy2019;
```



# 03

## 작업과정 - 데이터 수집

### 1. 삽입된 Data 확인

```
hive (happiness)> select * from happy2019;  
OK
```

1	Finland	7.769	1.34	1.587	0.986	0.596	0.393	0.153	
2	Denmark	7.6	1.383	1.573	0.996	0.592	0.41	0.252	
3	Norway	7.554	1.488	1.582	1.028	0.603	0.341	0.271	
4	Iceland	7.494	1.38	1.624	1.026	0.591	0.118	0.354	
5	Netherlands	7.488	1.396	1.522	0.999	0.557	0.298	0.322	
6	Switzerland	7.48	1.452	1.526	1.052	0.572	0.343	0.263	
7	Sweden	7.343	1.387	1.487	1.009	0.574	0.373	0.267	
8	New Zealand	7.307	1.303	1.557	1.026	0.585	0.38	0.33	
9	Canada	7.278	1.365	1.505	1.039	0.584	0.308	0.285	
10	Austria	7.246	1.376	1.475	1.016	0.532	0.226	0.244	
11	Australia	7.228	1.372	1.548	1.036	0.557	0.29	0.332	
12	Costa Rica	7.167	1.034	1.441	0.963	0.558	0.093	0.144	
13	Israel	7.139	1.276	1.455	1.029	0.371	0.082	0.261	
14	Luxembourg	7.09	1.609	1.479	1.012	0.526	0.316	0.194	
15	United Kingdom	7.054	1.333	1.538	0.996	0.45	0.278	0.348	
16	Ireland	7.021	1.499	1.553	0.999	0.516	0.31	0.298	
17	Germany	6.985	1.373	1.454	0.987	0.495	0.265	0.261	
18	Belgium	6.923	1.356	1.504	0.986	0.473	0.21	0.16	
19	United States	6.892	1.433	1.457	0.874	0.454	0.128	0.28	
20	Czech Republic	6.852	1.269	1.487	0.92	0.457	0.036	0.046	
21	United Arab Emirates		6.825	1.503	1.31	0.825	0.598	0.182	0.262
22	Malta	6.726	1.3	1.52	0.999	0.564	0.151	0.375	
23	Mexico	6.595	1.07	1.323	0.861	0.433	0.073	0.074	
24	France	6.592	1.324	1.472	1.045	0.436	0.183	0.111	
25	Taiwan	6.446	1.368	1.43	0.914	0.351	0.097	0.242	
26	Chile	6.444	1.159	1.369	0.92	0.357	0.056	0.187	
27	Guatemala		6.436	0.8	1.269	0.746	0.535	0.078	0.175
28	Saudi Arabia		6.375	1.403	1.357	0.795	0.439	0.132	0.08
29	Qatar	6.374	1.684	1.313	0.871	0.555	0.167	0.22	
30	Spain	6.354	1.286	1.484	1.062	0.362	0.079	0.153	

# 03

## 작업과정 - 데이터 처리(zeppelin)

### 1. 제플린 백그라운드에서 실행

```
[hadoop@dn01 ~]$ 'zeppelin.sh start &
```

### 1. 제플린 jdbc 설정 후 실행

jdbc %jdbc

Option

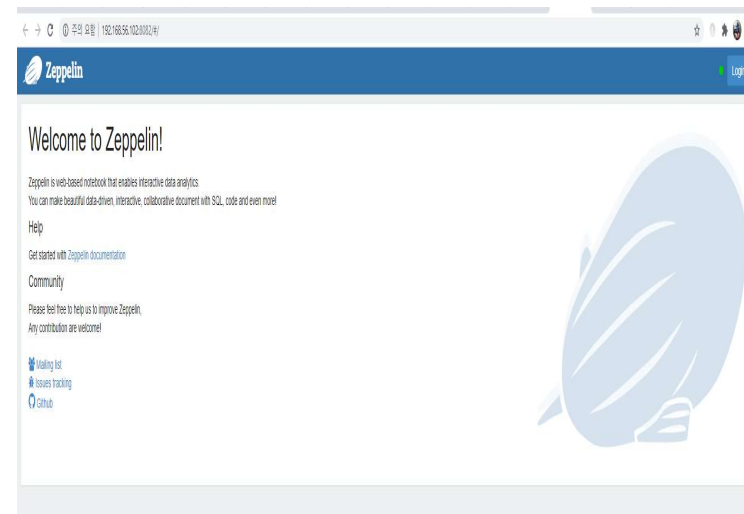
The interpreter will be instantiated [Globally](#) [in](#) [shared](#) [process](#)

☐ Connect to existing process

☐ Set permission

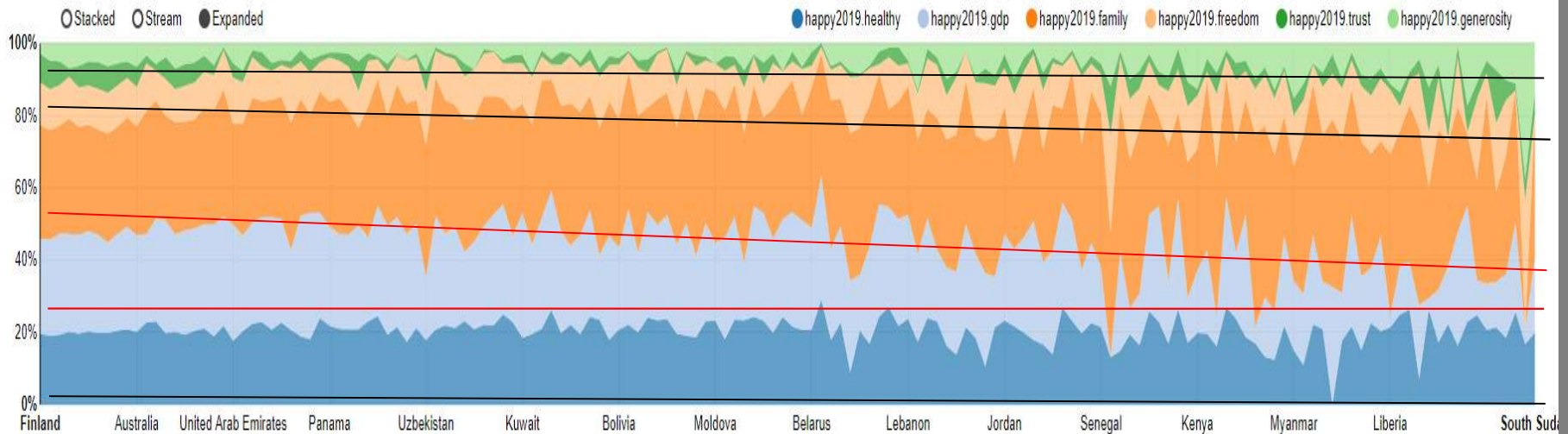
Properties

name	value
common.max_count	1000
default.completer.schemaFilters	
default.completer.ttlInSeconds	120
default.driver	org.apache.hive.jdbc.HiveDriver
default.password	***
default.precode	
default.splitQueries	false
default.statementPrecode	
default.uri	jdbc:hive2://192.168.56.102:10000
default.user	hive
zeppelin.jdbc.auth.type	
zeppelin.jdbc.concurrent.max_connection	10
zeppelin.jdbc.concurrent.use	true
zeppelin.jdbc.interpolation	false
zeppelin.jdbc.keytab.location	
zeppelin.jdbc.maxConnLifetime	-1
zeppelin.jdbc.maxRows	1000
zeppelin.jdbc.principal	



## 04 구현내역

가설 1: 행복도에 가장 영향을 끼치는 것은 자유일 것이다.



결론

그래프 내 선의 간격이 줄어 들수록, 행복도와 상관관계가 있다고 볼 수 있다.

행복도에 영향을 크게 미치는 항목은 GDP로 해석이 가능  
기대수명과 가족, 정부 부정부패는 영향이 없는 것은 아니지만 영향이 미미하다.

## 04 구현내역

- 2015-2019 GDP 변화 추이 쿼리문

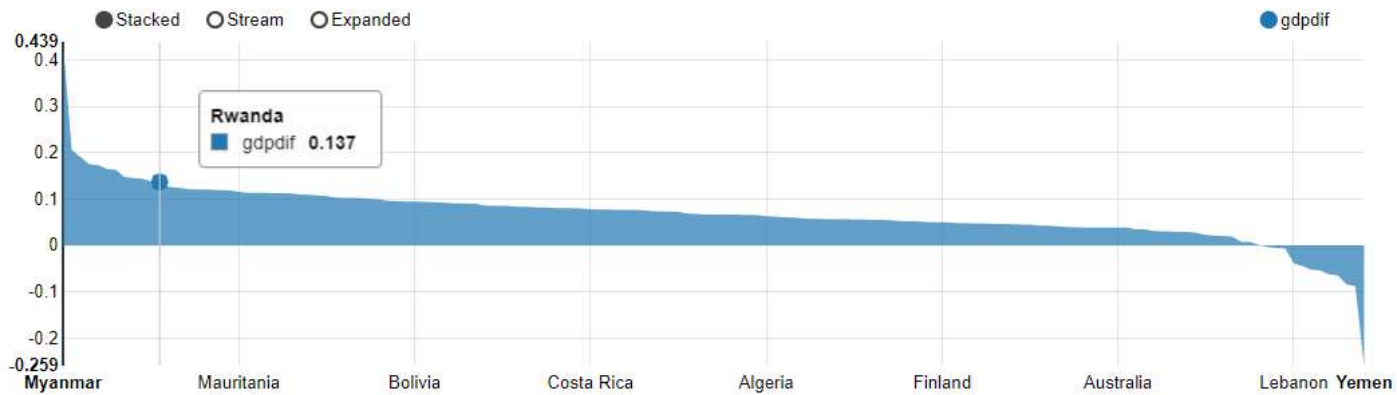
```
SELECT nine.country , (nine.gdp-fiv.gdp) as gdpdif  
FROM happy2019 as nine join happy2015 as fiv  
on nine.country=fiv.country  
ORDER BY gdpdif DESC
```

- 2015-2019 GDP 변화에 따른 행복도 변화

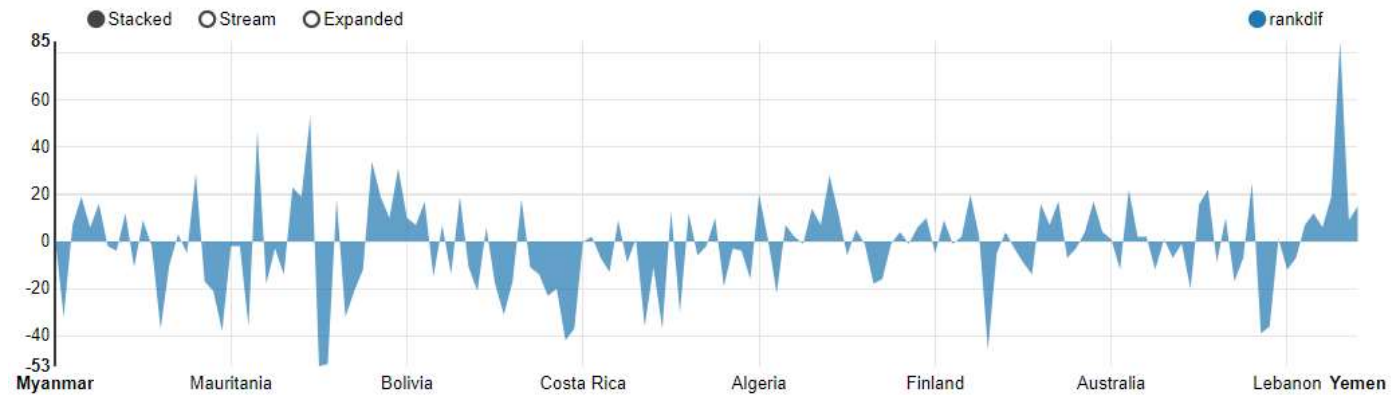
```
SELECT nine.country , ((nine.gdp-fiv.gdp)*1000) as gdpdif, (nine.hrank-fiv.hrank) as rankdif  
FROM happy2019 as nine join happy2015 as fiv  
on nine.country=fiv.country  
ORDER BY gdpdif DESC  
|
```

## 04 구현내역

가설 2: GDP의 상승은 행복지수의 상승으로 이어질 것이다.



- 2015-2019 GDP 변화 추이



- 2015-2019 GDP 변화에 따른 행복도 변화

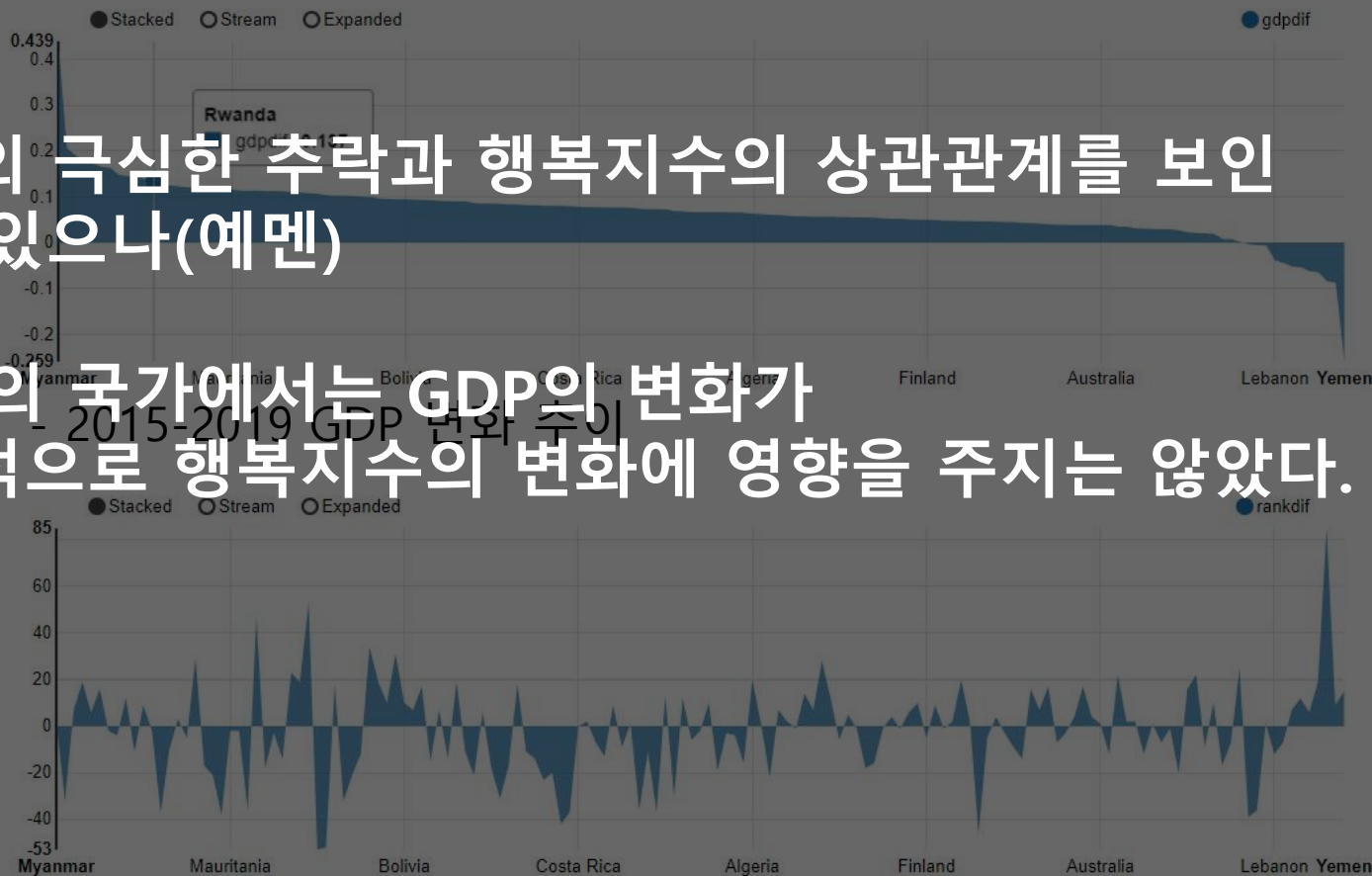
## 04 구현내역

가설 2: GDP의 상승은 행복지수의 상승으로 이어질 것이다.

결론:

GDP의 극심한 추락과 행복지수의 상관관계를 보인 예도 있으나(예멘)

그 외의 국가에서는 GDP의 변화가 즉각적으로 행복지수의 변화에 영향을 주지는 않았다.



- 2015-2019 GDP 변화에 따른 행복도 변화

**THANK  
YOU**

5조 김용선 문소희 박세영 박하얀 서영철 정아인