

2022.04

장경준

# Portfolio





# a Portfolio About

---

- About Me
- 데이터 분석 주제 선정
- 데이터 수집 및 정제
- 데이터 시각화
- 데이터 분석
- 결론

# About Me



## 데이터 엔지니어 지망 장경준

Python과 R을 기반으로 Data Engineering을 꿈꾸는 주니어 개발자 장경준 입니다.

“경험이 최고의 스승이다” 라는 좌우명을 바탕으로 항상 새로운 것을 도전하고 부딪히길 두려워하지 않았습니다. 대학교와 부산의 IT 전문 학원인 “IT Bank”에서 프로그래밍 기초인 C, Java, Python 및 네트워크와 리눅스 시스템을 배우며 흥미를 가지기 시작했고, 대학교 고학년 때 Kotlin 언어를 이용한 안드로이드 앱 개발, 해킹 실무 및 보안, 사이버 포렌식, 데이터 마이닝 등을 배우며 개발자에 대한 꿈을 확실히 했습니다.

대학교 졸업 이후, “빅데이터” 라는 정보가 가진 힘에 강렬한 이끌림을 느껴 Data Engineering에 대한 진로를 확고히 했고, 데이터에 대해 조금 더 공부를 하기 위해 “부산 더조은 직업전문학교” 에서 5개월간의 직업 훈련인 “빅데이터 개발자 양성” 과정을 수료했습니다.

“더조은 직업전문학교” 에서 HTML, JSP, Java, Spring Boot를 이용한 Web Front-End 개발 기초과정을 익혔고 Python과 R을 이용한 데이터 수집 처리 및 시각화, Hadoop Echo System을 통한 데이터 관리를 배웠고, TensorFlow등의 ML/DL 과정 또한 배웠습니다.

## EDUCATION/EXPERIENCE :

- 동서대학교 컴퓨터공학 학사 2014.03 – 2021.02
- 부산 더조은 직업 전문학교 빅데이터 전문가 양성(파이썬, R) 2021.03 – 2021.08

## Certificate :

- 정보처리기사
- 빅데이터 분석 기사 필합
- ADsP (데이터분석 준전문가)
- 리눅스 마스터 2급
- SQLD (SQL 개발자)
- 운전면허 1종 보통
- 네트워크 관리사 2급

## AWARDS :

- |    |                               |         |
|----|-------------------------------|---------|
| 대상 | 동서대학교 주관 교내<br>사이버포렌식/암호크랙 대회 | 2019.10 |
|----|-------------------------------|---------|

## CONTACT :

- |        |   |
|--------|---|
| Phone  | 010-6566-8843   |
| E-mail | egag8843@gmail.com  |
| Github | <a href="https://github.com/egag8843">https://github.com/egag8843</a> |

# 01

## 분석 주제 : 던파 복귀 유저의 복귀 원인

던전앤파이터를 플레이 하는 유저들은 저마다의 이유로 게임을 그만둡니다. 반복된 파밍에 지쳐서, 엔드급 스펙을 갖추고 더 이상 즐길 콘텐츠가 없어서, 애정 하던 무기를 강화하다 파괴되어서, 혹은 더 이상 재미가 없어서. 하지만, 던파 커뮤니티에는 이런 말이 있습니다.

“휴던은 있어도 탈던은 없다.”

그만큼, 던파에 대한 애정을 못 잊어 돌아오게 되는 복귀유저가 많습니다. 저는 유저들이 던파의 어떤 부분에 매료되어 다시 돌아오게 되는지 흥미를 느껴 던전앤파이터 복귀 유저에 대한 복귀 원인을 분석 주제로 정했습니다.



<출처 : 이코노믹 리뷰, 2018.07.14, <http://www.econovill.com/news/articleView.html?idxno=341601>>



<출처 : 던파 유튜버 '강개', 2018.10.17, [https://www.youtube.com/watch?v=E9T\\_0u9fAFM&t=5s](https://www.youtube.com/watch?v=E9T_0u9fAFM&t=5s)>

# 02

## 데이터 수집 및 정제 : Dataset 준비

던전앤파이터 복귀에 대한 관련 데이터 수집을 위해 데이터 수집 기법 중 하나인 “웹 크롤링”을 사용하여 Dataset 을 만들었습니다.

네이버 블로그에 “던파 복귀“ 라는 단어를 검색한 후, 검색 결과 등장하는 모든 블로그의 url을 순회하며 내용을 추출 하였습니다.

또한, 던파 복귀에 큰 영향이라 생각하는 추억, 추억에 대한 내용에 대해 데이터를 얻기 위해 유튜브 “던파 추억 ” 의 검색결과에 대한 제목과 조회수를 추출하여 Dataset을 만들었습니다.

### # 개발 환경

1. 운영체제 버전 : Windows 11 Pro
2. 사용 언어 : Python
3. Python Version : Python 3.7.5
4. 개발 IDE : Pycharm Community Edition 2022.1

네이버 블로그 검색결과로 생성된 데이터 중 명사만 추출하여 약 65000개의 Dataset을 생성하였고, 파이썬의 라이브러리를 이용하여 시각화 하였습니다.

유튜브 검색결과로 생성된 데이터는 약 500개 이며, 엑셀을 통해 데이터 정제 및 시각화 하였습니다.

- Dataset 1 : [df\\_blog\\_crawling.xlsx](#)
- Dataset 2 : [youtube\\_crawling.xlsx](#)
- 유튜브 검색결과 시각화 파일 : [youtube\\_visualization.xlsx](#)

## 2-1. Dataset 생성 과정

### Naver Blog 검색 결과 크롤링

- 던전앤파이터를 플레이 했던 유저들이 다시 복귀하게 된 이유를 알기 위해 네이버 블로그에 "던파 복귀" 검색 후 콘텐츠 크롤링.
- Selenium 라이브러리를 사용하여 네이버 블로그 중 "던파 복귀" 검색 결과 블로그들의 url 추출.
- 파이썬 파일 : dfBlogCrawling.py

```

1  import ...
13  |
14  # xpath 사용 시 등장하는 deprecation warning 해제
15  warnings.filterwarnings(action="ignore")
16
17  """
18  ##### 데이터 수집 - 웹 크롤링 #####
19  """
20
21  path = "chromedriver.exe" # 크롬 웹 드라이버 경로 지정
22
23  options = webdriver.ChromeOptions()
24  options.add_argument('headless') # 백그라운드 실행
25
26  driver = webdriver.Chrome(path, chrome_options=options) # 드라이버 경로 지정
27  url_list = [] # 블로그 url 저장
28  content_list = "" # 블로그 콘텐츠 누적 변수
29  text = "던파 복귀" # 검색어
30  #text = "던파 BGM"
31
32  # 1~500 페이지 블로그의 링크 추출
33  for i in tqdm(range(1, 501), desc = '블로그 링크 추출'):
34      url = 'https://section.blog.naver.com/Search/Post.nhn?pageNo=' + str(i) + '&rangeType=ALL&orderBy=sim&keyword=' + text
35      driver.get(url)
36      time.sleep(0.5) # 오류 방지
37
38      for j in range(1, 3):
39          titles = driver.find_element_by_xpath('/html/body/ui-view/div/main/div/div/section/div[2]/div['+str(j)+']/div/div[1]/div[1]/a[1]')
40          title = titles.get_attribute('href')
41          url_list.append(title)
42
43
44  print("\n블로그 url 수집 완료, 해당 url 데이터 크롤링\n")

```

## 2-1. Dataset 생성 과정

### Naver Blog 검색 결과 크롤링

#### # 데이터 수집

- 수집한 url들을 차례대로 순회하며 블로그 내용에 해당하는 html 태그를 선택하여 크롤링하여 리스트에 저장.

#### # 데이터 정제

- 형태소 분석 라이브러리인 Okt를 사용하여 추출한 내용 중 명사만 추출.
- 추출한 내용 중 불용어 제거. 기본 조사 및 데이터 분석에 악영향을 미치는 단어 제거.

```

46 # 링크 추출된 블로그들 순회하며 블로그 내의 내용 크롤링
47 try:
48     for url in tqdm(url_list, desc='블로그 내용 크롤링'):
49         driver.get(url)
50         driver.switch_to.frame("mainFrame")
51         overlays = ".se-component.se-text.se-l-default" # 내용 크롤링
52         contents = driver.find_elements_by_css_selector(overlays)
53
54         for content in contents:
55             content_list = content_list + content.text # 각 블로그 내용 누적
56 except Exception as e:
57     print("예외 발생 : ", e)
58
59 """
60 ##### 크롤링된 데이터 정제 - 명사 추출 및 불용어 제거 #####
61 형태소 분석기 Okt를 사용하여 수집한 데이터 중 명사만 추출
62 기본 조사 및 데이터 분석 시 의미 없는 데이터 제거
63 Ex. 던파, 1호기, 몇 일자 등 검색 시 "던파 북귀"에 관한 검색 시 불가피하게 등장하는 단어들 제거
64 """
65
66 # 형태소 분석기 Okt 사용
67 okt = Okt()
68 myList = okt.pos(content_list, norm=True, stem=True) # 모든 형태소 추출
69 myList_filter = [x for x, y in myList if y in ["Noun"]] # 추출된 값 중 명사만 추출
70 Okt_before = Text(myList_filter, name="Okt")
71
72 # 데이터 정제 전 데이터 개수
73 freq = Counter(Okt_before).most_common()
74 print("데이터 정제 전 데이터 개수 : ", len(Okt_before))
75
76 with open('data/stopwords.txt', 'r') as f:
77     list_file = f.readlines()
78
79 stopwords = list_file[0].split(",") # 콤마로 구분
80 remove_wordList = [x for x in Okt_before if x not in stopwords]
81 Okt_after = Text(remove_wordList, name="Okt")
82 freq_after = Counter(Okt_after).most_common()

```

## 2-1. Dataset 생성 과정

### Naver Blog 검색 결과 크롤링

#### # 데이터 시각화

- Matplotlib 라이브러리를 사용하여 명사 별 출현 빈도수에 대해 막대그래프로 표현.
- Wordcloud 라이브러리를 사용하여 워드클라우드 출력.

#### # 마무리

- 정제가 마무리된 데이터를 Excel 파일로 저장.

```

87 """
88 ##### 데이터 시각화 #####
89 """
90 # 그래프 한글 출력 인코딩
91 font_location = "c:/Windows/Fonts/malgun.ttf"
92 font_name = font_manager.FontProperties(fname=font_location).get_name()
93 rc('font', family=font_name)
94
95 # 그래프 x, y 라벨 설정
96 plt.xlabel("단파 복귀 관련 명사")
97 plt.ylabel("빈도수")
98
99 # 그래프 x, y 값을 설정
100 wordInfo = dict()
101 for tags, counts in Okt_after.vocab().most_common(50):
102     if len(str(tags)) > 1:
103         wordInfo[tags] = counts
104
105 values = sorted(wordInfo.values(), reverse=True)
106 keys = sorted(wordInfo, key=wordInfo.get, reverse=True)
107
108 # matplotlib 그래프 값 설정
109 plt.bar(range(len(wordInfo)), values, align='center')
110 plt.xticks(range(len(wordInfo)), list(keys), rotation='100')
111 plt.title('단파 복귀 검색어에 대한 명사 별 빈도수')
112 plt.show()
113
114 # wordcloud 출력
115 wc = WordCloud(width = 1000, height = 1000, background_color = "white", colormap = 'autumn', font_path = font_location, max_words = 50)
116 plt.imshow(wc.generate_from_frequencies(Okt_after.vocab()))
117 plt.axis("off")
118 plt.show()
119
120 # Dataset 내보내기
121 dataset = pd.DataFrame(list(Okt_after), columns=['추출한 명사'])
122 dataset.to_excel('df_blog_crawling.xlsx', encoding='', index=False)
123

```



## 2-2. Dataset 생성 과정 Youtube 검색 결과 크롤링

### # 데이터 수집, 정제 및 시각화

- BeautifulSoup, Selenium 라이브러리를 사용해 유튜브 “던파 추억” 검색 결과 중 제목과 조회수 추출.
- 데이터 추출을 통해 생성된 Excel 파일을 이용해 데이터 정제 및 시각화.
- 파이썬 파일 : dfYoutubeCrawling.py

```

1  import ...
10
11  # tag_name 사용 시 등장하는 deprecation warning 해제
12  warnings.filterwarnings(action='ignore')
13
14  """
15  ##### 데이터 수집 - 웹 크롤링 #####
16  """
17
18  text = "던파" 추억 # 검색어, 명확한 유튜브 검색을 위해 ""로 묶음.
19  path = "chromedriver.exe" # 크롬 웹 드라이버 경로 지정
20  url = 'https://www.youtube.com/results?search_query=' + text # 유튜브 url 지정
21
22  options = webdriver.ChromeOptions()
23  options.add_argument('headless') # 백그라운드 실행
24
25  driver = webdriver.Chrome(path, chrome_options=options)
26  driver.get(url)
27
28  # 페이지 정보가 있는 body 태그 추출
29  body = driver.find_element_by_tag_name('body')
30
31  for i in range(1,300): # body 태그를 통해 페이지를 내리며 검색
32      body.send_keys(Keys.PAGE_DOWN)
33      time.sleep(0.3) # 오류방지
34
35  # soup를 통해 데이터 가져오기
36  soup = bs(driver.page_source, 'lxml')
37
38  # 제목, 주소, 조회수가 포함된 a#video-title 추출
39  title = soup.select('a#video-title')
40  view = soup.select('a#video-title')
41
42  title_list = []
43  view_list = []
44
45  for i in range(len(title)): # 제목과 조회수를 리스트에 저장
46      title_list.append(title[i].text.strip())
47
48  # 'aria-label' 속성에서 문자열 나눈 후 마지막 값 사용, 정수형 변환을 위해 텍스트 자름.
49  view_list.append(view[i].get('aria-label').split()[-1].strip('회분').replace(',',''))
50
51  # 추출한 제목과 조회수 딕셔너리에 저장
52  youtubeDic = {
53      '제목':title_list,
54      '조회수':view_list
55  }
56
57  #Dataset 내보내기
58  df = pd.DataFrame(youtubeDic)
59  df.to_excel('youtube_crawling.xlsx', encoding='', index=False)
60

```

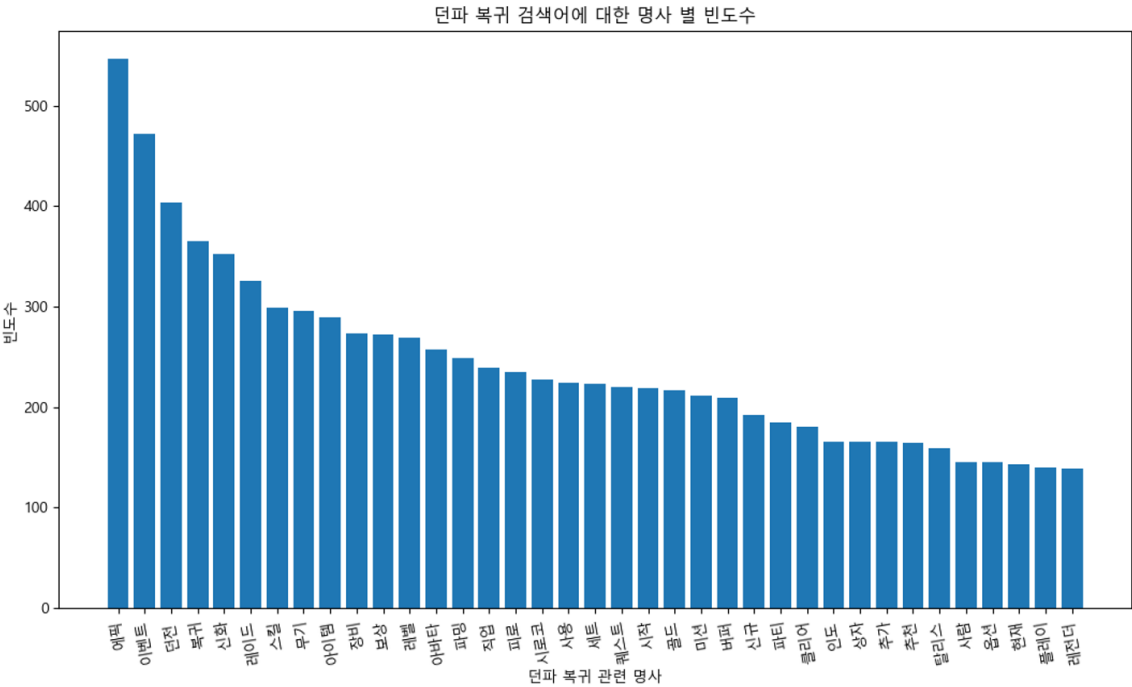
# 03

## 데이터 분석 : 시각화 된 자료를 통한 분석

좌상단의 그래프는 네이버 블로그 “던파 복귀 ” 에 관한 검색 결과들 중 명사만 추출하여 명사 별 빈도수를 나타낸 그래프입니다.

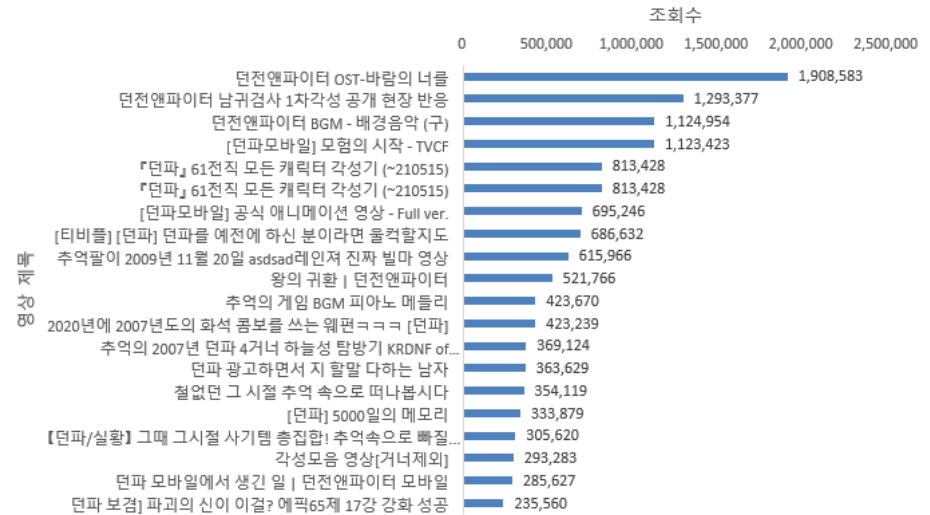
크롤링을 통해 도출한 명사 중 가장 많은 빈도수를 차지한 단어는 “에픽과 이벤트 ” 입니다. 복귀 유저가 더욱 편하고 즐거운 게임 플레이를 할 수 있도록 지원해주는 던파의 복귀 유저 이벤트와 많은 보상, 특히 에픽 등급 아이템과 신화 등급 아이템을 주는 이벤트가 복귀의 가장 큰 원인이었습니다.

그 뒤로 던전, 신화, 아이템, 파밍, 무기, 레벨 등 자신의 캐릭터가 강해지는 이른바 “스펙 업 ” 이 주를 이루었고, 레이드, 시로코와 같이 근 3년간 던파에서 이루어져 왔던 대규모 업데이트의 단어도 많이 추출되었습니다.



1	제목	조회수
2	던전앤파이터 OST-바람의 너를	1,908,583
3	던전앤파이터 남귀검사 1차각성 공개 현장 반응	1,293,377
4	던전앤파이터 BGM - 배경음악 (구)	1,124,954
5	[던파모바일] 모험의 시작 - TVCF	1,123,423
6	『던파』 61전직 모든 캐릭터 각성기 (~210515)	813,428
7	『던파』 61전직 모든 캐릭터 각성기 (~210515)	813,428
8	[던파모바일] 공식 애니메이션 영상 - Full ver.	695,246
9	[티비플] [던파] 던파를 예전에 하신 분이라면 울컥할지도	686,632
10	추억팔이 2009년 11월 20일 asdsad레인저 진짜 빌마 영상	615,966
11	왕의 귀환   던전앤파이터	521,766
12	추억의 게임 BGM 피아노 메들리	423,670
13	2020년에 2007년도의 화석 콤보를 쓰는 웨펀ㅋㅋㅋ [던파]	423,239
14	추억의 2007년 던파 4거너 하늘성 탐방기 KRDNF of memories 2007 KRDNF Season1	369,124
15	던파 광고하면서 지 할말 다하는 남자	363,629
16	철없던 그 시절 추억 속으로 떠나봅시다	354,119
17	[던파] 5000일의 메모리	333,879
18	【던파/실황】 그때 그시절 사기템 총집합! 추억속으로 빠질 준비 되었나요?!	305,620
19	각성모음 영상[거너제외]	293,283
20	던파 모바일에서 생긴 일   던전앤파이터 모바일	285,627
21	던파 보검] 파괴의 신이 이걸? 에픽65제 17강 강화 성공	235,560

"던파 추억" 유튜브 검색 결과 조회수 상위 20개의 영상 제목



&lt;"던파 추억" 관련 Youtube 검색 결과 시각화 엑셀 자료&gt;

Youtube 크롤링 결과, 추억에 관련된 영상들이 500개 가량 추출되었습니다. 상위 조회수의 경우, 10만회 이상 조회된 영상도 많았고 가장 높은 조회수를 기록한 영상은 과거 던파의 로그인 화면을 책임졌던 추억의 OST "바람의 너를"이었습니다.

상위 조회수의 영상들을 살펴본 바, 대전이 이전의 던전들과 던전들의 배경음악, 1차 각성 퀘스트의 진행 던전이었던 고대던전의 영상들이 주를 이루었습니다. 최근 출시된 던파 모바일 또한 "던파 클래식"이라는 호평을 받으며 추억을 회상하는 많은 유저들을 불러모았습니다.

이렇듯, 던파 추억과 관련된 영상도 많고 그 추억을 되살리기 위해 다시 던파로 돌아오는 유저들, 던파 복귀 이벤트 보상에 다시 돌아오는 유저들, 대규모 업데이트에 흥미를 느껴 돌아오는, 절대 등한시 할 수 없는 복귀 유저들이 꾸준히 있었습니다.

## 04

결론 :  
분석을 통한 결론

가장 최근 대규모 업데이트인 '더 넥스트 저니', 110 만렙 확장과 동시에 새로운 레벨대의 에픽 장비, 신규 던전, 캐릭터의 밸런스 패치와 개편과 그와 동시에 진행했던 이벤트로 인해 PC방 점유율이 급상승한 것을 확인할 수 있습니다.

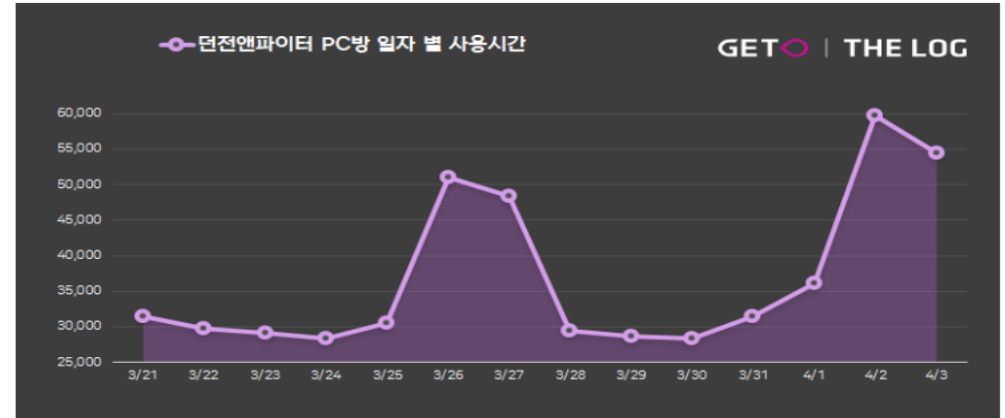
또한, 작년에 진행됐던 "혼돈의 오즈마" 업데이트 당시에도 가파른 접속률 상승세를 이루어 냈습니다. 시로코 레이드 업데이트 이후로 유저들이 가장 기다렸을 새로운 레이드인 오즈마 레이드. 저 또한 당시 많은 기대를 하며 설레었던 경험이 있습니다.

이를 통해, 대규모 업데이트 때 많은 유저들이 복귀하고, 복귀 시 도태되지 않도록 진행되는 복귀 이벤트, 그리고 기존 유저들을 위한 이벤트 역시 복귀 유저들도 마음껏 누릴 수 있어 이벤트의 내용도 복귀에 있어 매우 중요한 요소라는 결론을 도출할 수 있습니다.

## &lt;이주의 게임&gt;

이주의 게임으로 던전애파이터가 선정됐다.

던전애파이터는 지난 3월 3주차 시즌8 '더 넥스트 저니 (THE NEXT JOURNEY)' 돌입 이후 꾸준한 상승세와 함께 모바일 버전 출시의 화제성이 더해지며 PC방 사용량이 증가했다.



<던파 대규모 업데이트 당시 PC방 점유율>

출처 : 더로그, <https://www.thelog.co.kr/report/reportList.do> >

그 외에도 복귀 유저들의 복귀한 계기는 윤명진 총괄 디렉터님의 복귀 이후의 행보, 시너지 캐릭터와 딜러 캐릭터의 구분을 없애는 던파의 역대급 상향 평준화 밸런스 패치로 많은 호응을 얻었으며, 타 RPG게임과 달리 많은 극찬을 받았던 던파 간담회의 내용, 연례 행사로 진행되는 겨울 던파 페스티벌의 행사 내용 또한 던파를 돌아섰던 유저들을 다시 불러올 수 있는 충분한 요소라고 사료됩니다.

그리고 가장 최근 출시한 던파 모바일도 과거의 향수를 불러 일으키며 PC 던파 접속률 상승에도 좋은 영향을 미칠 것입니다.

2022.04

장경준

# Thank You

