

알고타자! 믿고타자! 같이타자!
오늘의 항공사 예측분석 서비스

ON:AIR

A203 팀장 박지우 팀원 김채은 김현욱 나승호 안수빈



CONTENTS

01 프로젝트 소개

1. 요구사항 및 니즈
2. 솔루션
3. 기존 서비스와의 차별점

02 프로젝트 운영

1. 형상관리
2. 아키텍처

03 빅데이터 활용

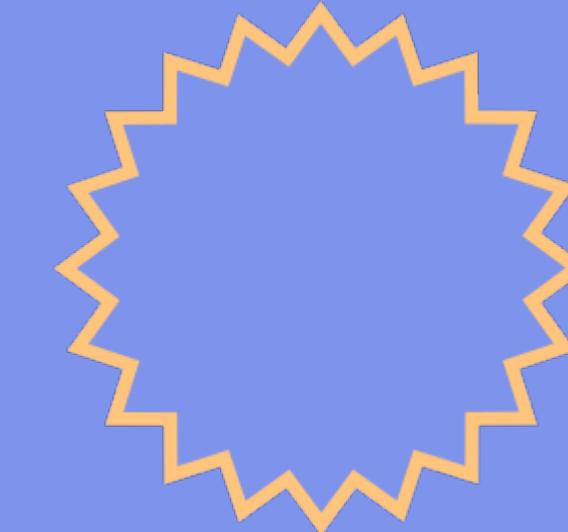
1. 항공 운항 데이터 통계
2. 출/도착 지역 예측 분석
3. 리뷰(자연어) 감성 분석

04 웹 시연

05 확장 가능성

01

프로젝트 소개



01. 요구사항 및 니즈
02. 기존 서비스와의 차별점

요구사항 및 니즈

다양한 항공사 자연의 원인을 분석해서 미리 예측할 수 없을까?



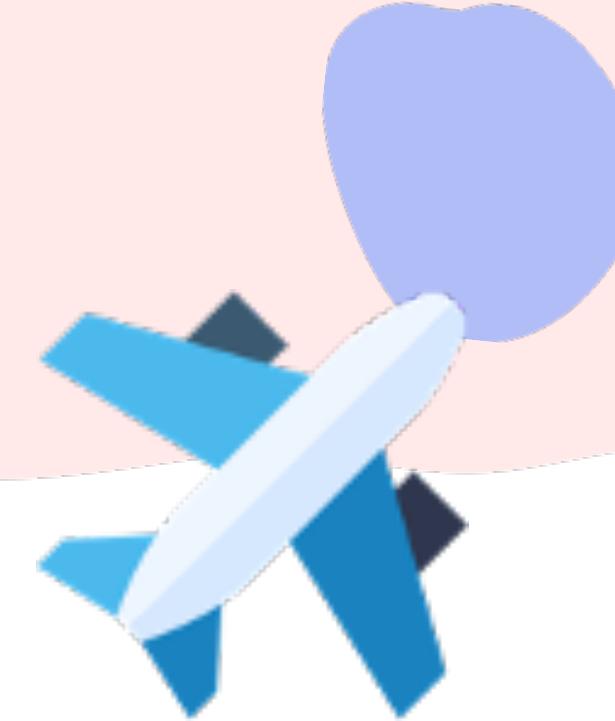
예비 여행객

여행 계획을 세울 때
자세한 항공사 정보를
얻고 싶은 여행객



항공사 이용객

항공사에 대한
솔직한 리뷰를
나누고 싶은 기 이용객



항공사

자사에 대한
솔직한 평가를
알고 싶은 항공사

솔루션

폭 넓은 항공
정보 제공



“
항공사 리포팅 서비스와
사용자의 리뷰
모두를 한 번에
확인할 수 있는 서비스
”

기존 서비스와의 차별점

On : Air

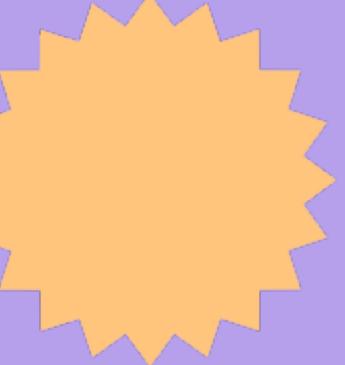
트립 어드바이저

구글 플라이트



항공사 기본 정보 제공	0	0	0
항공사 분석 리포트 제공	0	X	△
리뷰	0	0	X
지연 정보 및 사유	0	X	△

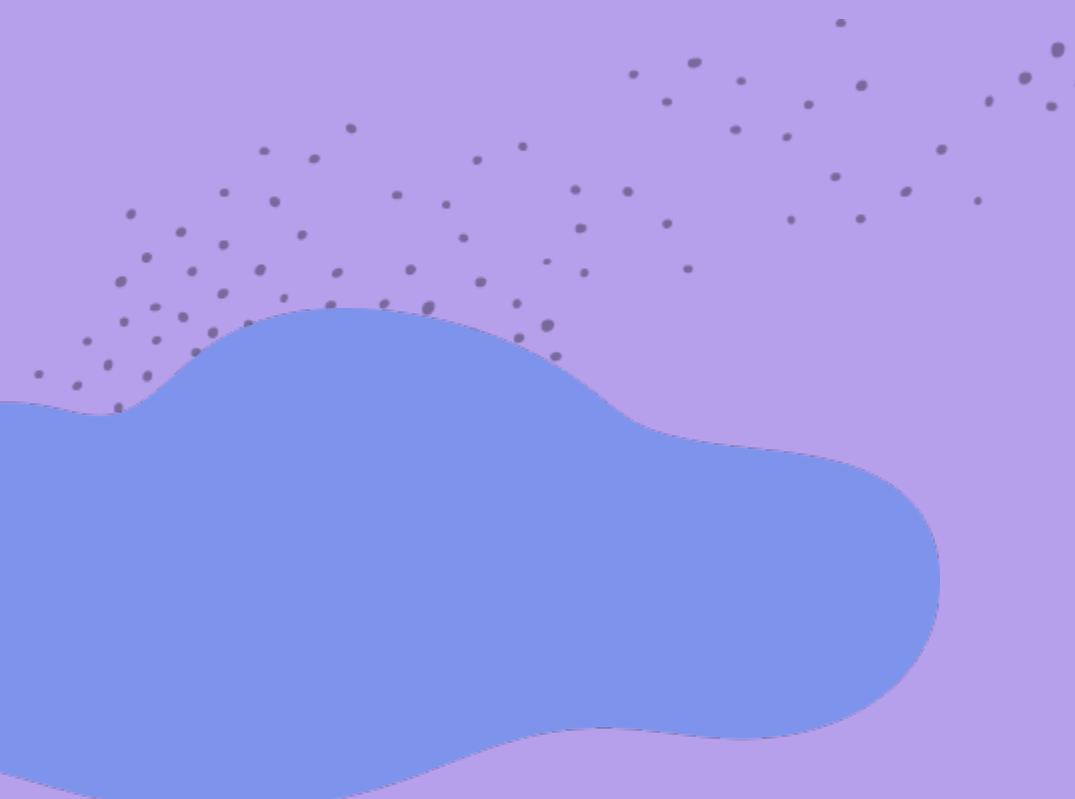
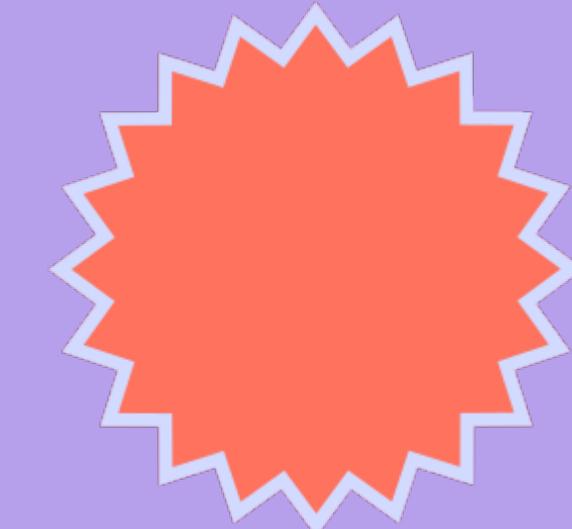
02



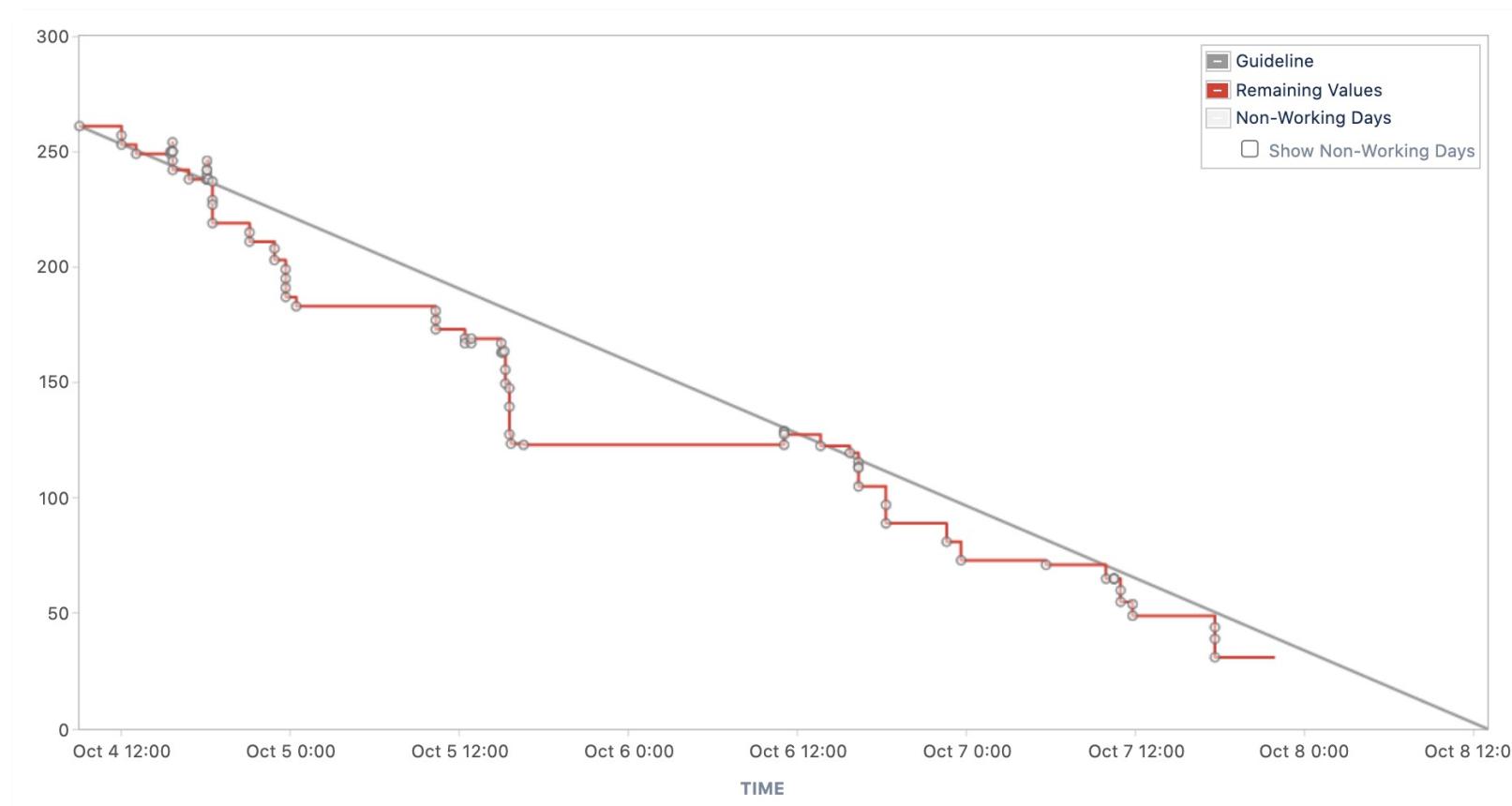
프로젝트 운영



- 01. 형상관리
- 02. 아키텍처



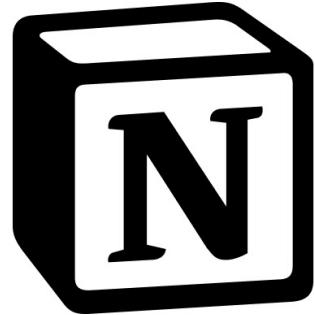
형상관리



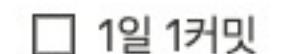
Commit statistics for master Aug 30 - Oct 07

Excluding merge commits. Limited to 2,000 commits.

- Total: **522 commits**
- Average per day: **13.4 commits**
- Authors: **7**



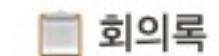
운영



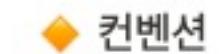
1일 1커밋



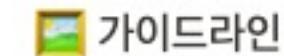
Webex On-Time



회의록



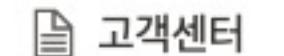
컨벤션



가이드라인



Clean Coding



고객센터



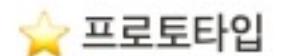
설계



디자인



와이어프레임



프로토타입



API(/api/v1)



Data Set



ERD



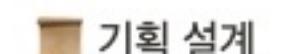
개발



기획



아이디어



기획 설계



Tech Stack



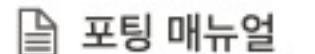
폴더 구조



스토리보드

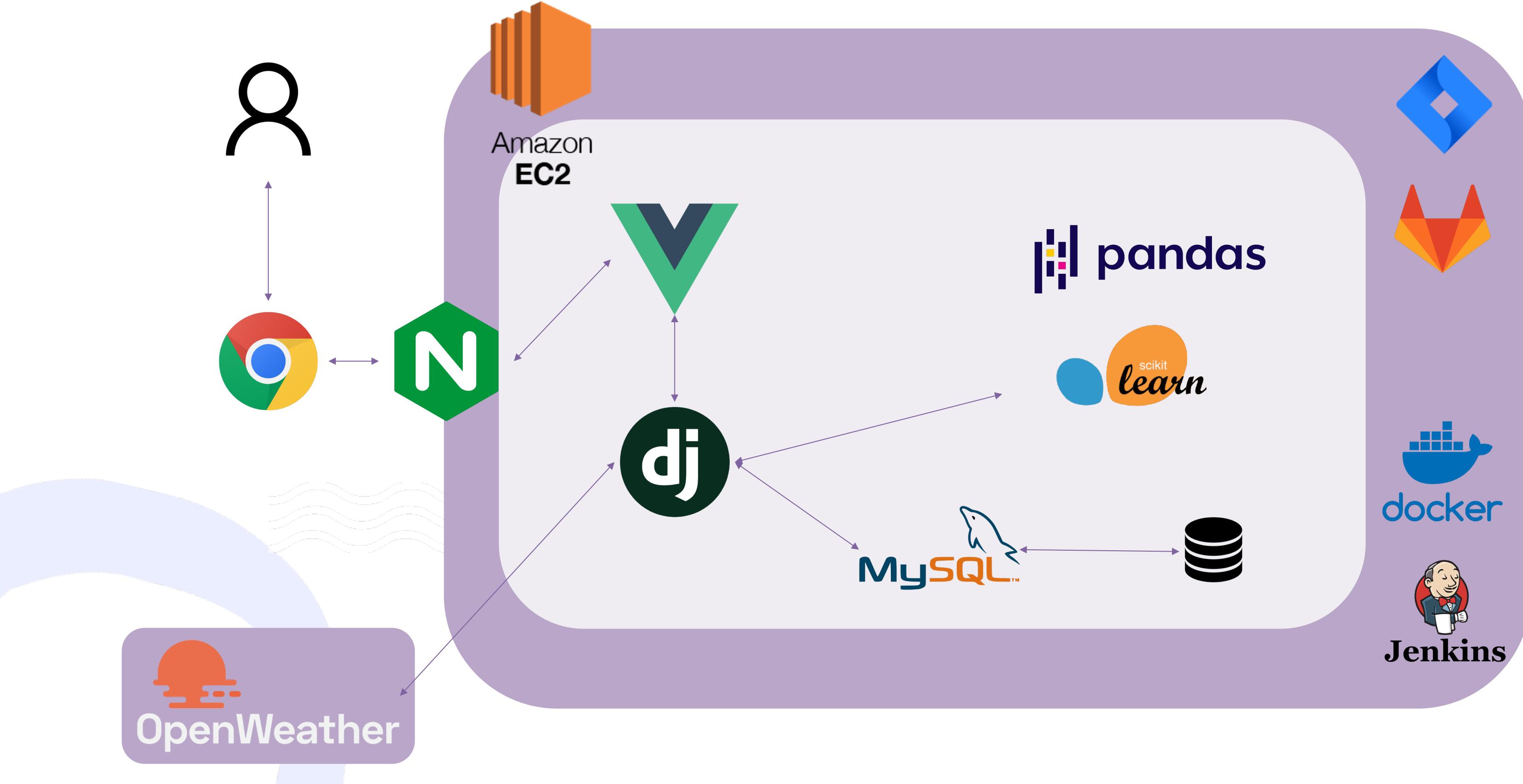


데이터 활용



포팅 매뉴얼

아키텍처

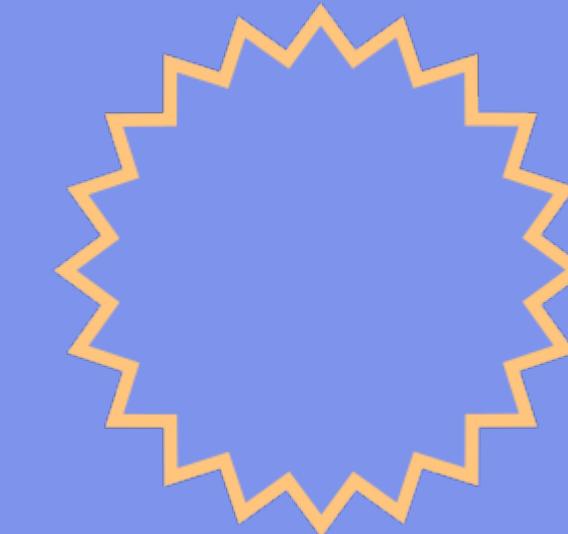


03

빅데이터 활용



01. 항공 운항 데이터 통계
02. 항공 지연 예측 분석
03. 항공사 리뷰 감성 분석



어떤 통계를 낼 것인가?

01

항공사 총 지연 및 결항 통계 (실적)

각 항공사의 모든 목적지에 대한 총합 통계 데이터 시각화
총 운항횟수 / 지연률 / 결항률 / 평균 지연시간

02

항공사의 목적지별 지연 통계 (이용객용)

이용자가 검색한 목적지에 대한 통계 데이터 시각화
총 운항횟수 / 지연률 / 평균 지연시간

03

지연 사유 분석

항공사의 모든 목적지에 대한 지연 사유 분포 시각화
월별 평균 출발 지연 시간 시각화
개별 목적지에 대한 지연 사유 분포 및 평균 지연 시간 시각화

04

월별 이용객 통계

각 항공사별 월별 이용객 추이 시각화
머신러닝에 활용되는 이용객 데이터 참고 자료로 제시

항공 운항 데이터 통계



인천공항 출/도착 데이터 (2017.01 ~ 2021.08)

- <https://www.airportal.go.kr/> (한국항공협회 공공데이터)



인천공항 항공사별 월별 이용객 데이터 (2017.01 ~ 2021.08)

- <https://www.airport.kr/> (인천국제공항공사)
- 활용 데이터 : 항공사 / 월별 / 여객수

항공 운항 데이터 통계



항공사	편명	목적지	계획	예상	출발	구분	현황
티웨이항공	TW337	GUM(괌)	06:55	06:55	07:12	여객	출발
비엣제트 항공	VJ879	DAD(다낭)	07:00	08:10	08:29	여객	지연
진에어	LJ201	NRT(나리타)	07:10	07:10	07:24	여객	출발
제주항공	7C1304	KIX(간사이)	07:10	07:10	07:21	여객	출발
이스타항공	ZE563	CXR(깜라안(ня짱))	07:15	10:30	10:49	여객	지연
에어서울	RS515	DAD(다낭)	07:15	07:15	07:35	여객	출발
비엣제트 항공	VJ925	HPH(카트비)	07:15	07:15	07:26	여객	출발
아시아나항공	OZ731	SGN(떤선녓엇(호찌민))	07:20	07:40	08:08	여객	출발
진에어	LJ077	DAD(다낭)	07:20	07:20	08:03	여객	출발
티웨이항공	TW145	KLO(칼리보)	07:25	07:25	07:57	여객	출발
티웨이항공	TW669	RMQ(타이중)	07:25	08:05	08:34	여객	지연
진에어	LJ021	CEB()	07:30	07:30	07:53	여객	출발
티웨이항공	TW605	TNA(지난)	07:30	07:30	07:43	여객	출발
대한항공	KE529	FRA(프랑크푸르트)	07:30	07:30	07:33	화물	출발

total	under_60	under_30	over_60	delay_rate	delay_time	airline	arrival
87	0	82.56	17.44	17.24	14	티웨이항공	HKG(홍콩)
31	0	96.77	3.23	3.23	2	대한항공	HGH(항저우)
1261	0.32	94.83	4.85	5.15	5	대한항공	RGN(양곤)
377	0	93.35	6.65	6.63	7	진에어	HNL(호놀룰루)
1	0	100	0	0	0	아시아나항공	VIE(빈)
2210	0.09	93.97	5.94	6.02	6	제주항공	WEH(위해)
1048	0.1	95.8	4.1	4.2	4	대한항공	XMN(사먼)
1	0	100	0	0	0	델타항공	JFK(뉴욕)
1514	0.33	96.49	3.18	3.5	3	진에어	KKJ(기타큐슈)
568	0	95.77	4.23	4.23	5	대한항공	AOJ(아오모리)
838	0.48	89.01	10.51	10.98	10	에어서울	HKG(홍콩)
1509	0.47	94.08	5.45	5.9	6	대한항공	DLC(대련)
1160	0.34	94.22	5.43	5.78	5	대한항공	XIY(시안)
1208	0.41	90.98	8.61	9.02	9	아시아나항공	NKG(남경)
2268	0.31	96.06	3.63	3.92	4	대한항공	CTS(삿포로)
2807	0.46	88.77	10.77	11.22	11	아시아나항공	MNL(마닐라)
61	0	90.16	9.84	9.84	9	아시아나항공	HAK(하이커우)
1	0	100	0	0	0	제주항공	PEK(베이징)
1063	0.19	95.35	4.46	4.61	4	대한항공	OKA(오키나와)
91	0	100	0	0	0	중국동방항공	YNJ(연길)

원본데이터를 항공사의 목적지별 통계용 데이터로 전처리

Machine Learning은 어디에 쓰였는가?

기상 변수 및 이용객 변수로 각 항공사의 목적지별 지연률 예측

01

기상데이터로 지연률 예측

2017~2021 인천공항의 기상 데이터와 당시 지연 여부
현재 기상 데이터에서의 지연률 계산

02

월별 이용객 수로 지연률 예측

2017~2021 항공사별 이용객 수와 당시 지연 여부
이용객 수 추이를 통해 이용객 수를 예측하고
이에 대한 지연률 계산

03

두 가지 변수 모두로 지연률 예측

기상 데이터와 월별 이용객수를 모두 포함하여 지연률
예측



원핫인코딩

로지스틱 회귀 모델

정확도 평균
92%

항공 운항지연분석 데이터



유료데이터



현재 기상 & 2017~2021 인천공항의 일일 기상 데이터

- <https://home.openweathermap.org/> (OpenWeather)
- 활용 데이터: Weather Conditions (상세 관측 정보)
- Ex. 맑음, 비, 눈, 태풍 등



인천공항 월별 이용객 데이터 (2017.01 ~ 2021.08)

- <https://www.airportal.go.kr/> (한국항공협회 공공데이터)
- 활용 데이터 : 항공사 / 월별 / 여객수

리뷰 분석은 어떻게 이루어졌는가?

01

자연어 처리 학습

토크나이징을 위한 형태소 분석과 품사 태깅을 지원하는
pykomoran 라이브러리 학습

02

리뷰 TOP5 키워드 도출

리뷰 텍스트에서 언급이 많은 키워드 5개 도출

03

리뷰 긍부정 비율 분석

리뷰 텍스트의 감성을
긍정어 - 부정어로 파악하여 비율 시각화

04

워드 클라우드 시각화

리뷰 텍스트에서 상위 100개 단어 중 빈도수 순으로
불용어 제외하여 워드클라우드로 시각화

리뷰 분석 데이터



리뷰 초기 데이터

- <https://www.tripadvisor.com/Airlines> (Tripadvisor)
- 활용 데이터: 리뷰 내용 (작성자 및 개인정보 일체 X)
- 초기 데이터 분석 실습 및 결과 도출을 위한 단순 벌크 데이터로 활용

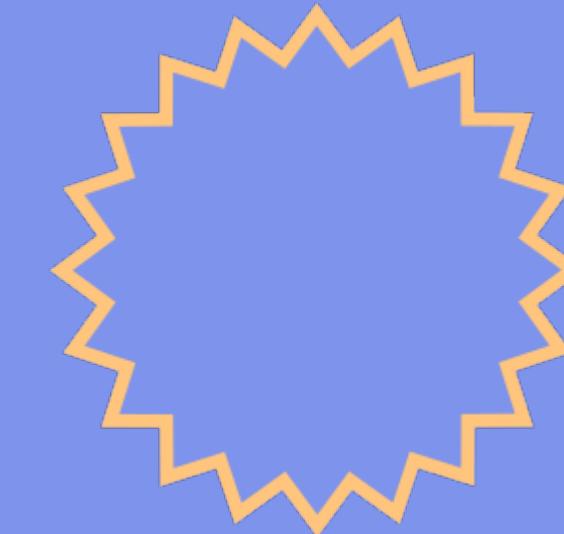


리뷰 사용자 작성 데이터

- 커뮤니티 서비스를 겸하고 있어 사용자가 가입 후 작성한 리뷰가 분석에 실시간 포함됩니다.

04

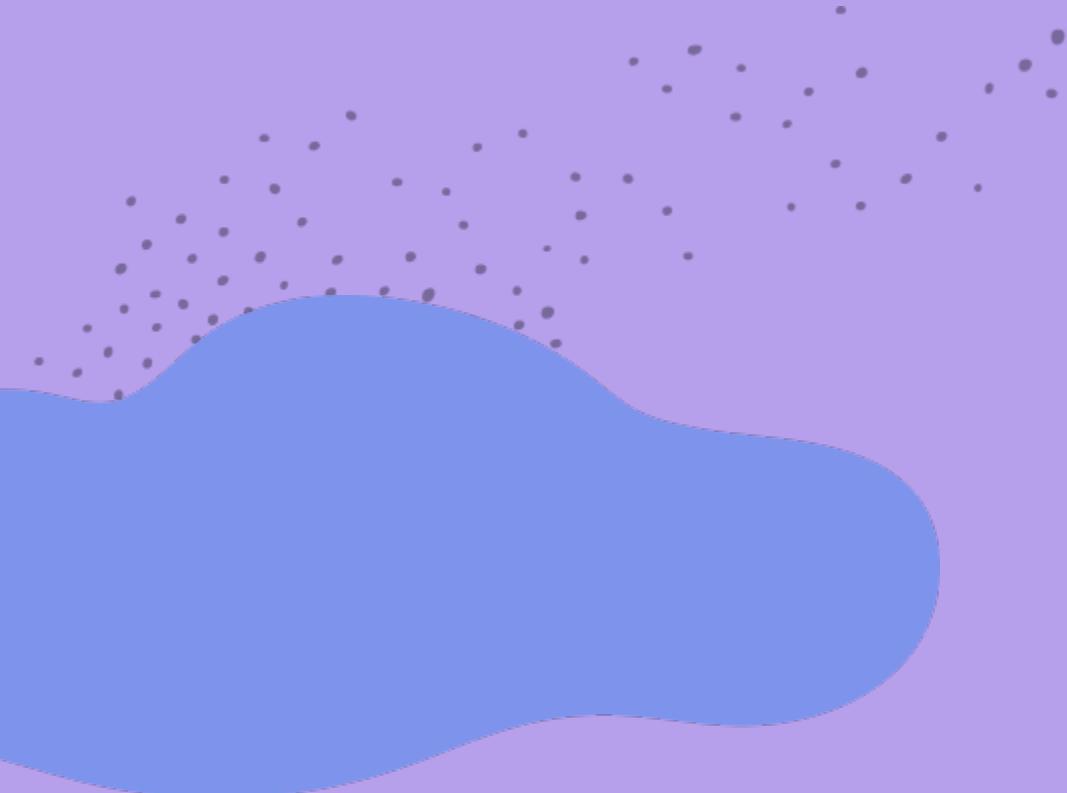
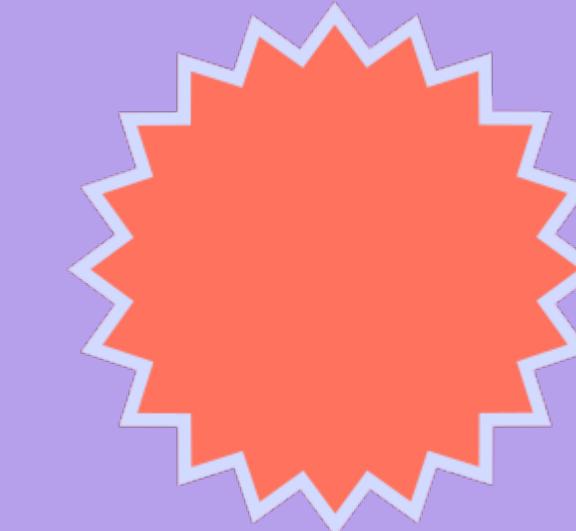
웹 시연



어디로 떠나볼까요?

05

확장 가능성



고도화 전략



“
인천공항 2017 -2021 출/도착 자료로 시작
차후 국내 주요공항 더 나아가 국제공항까지 데이터 규모를 늘려 서비스 확장 가능성”

Q & A



THANK YOU

발표를 들어주셔서 감사합니다 :)



밑으로
안씀니다



MySQL 연결

EC2 설정 및 Hadoop 설치

정의

데이터 분석 계획

Docker / Jenkins

배포

Test Scenario

기획

데이터 분석 과정

개발

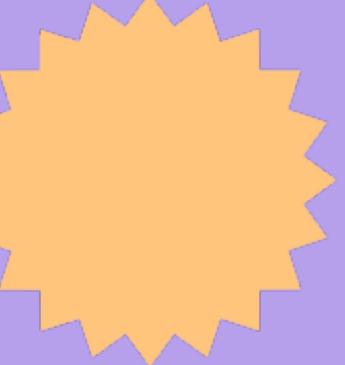
머신러닝 과정

운영

발표 문서

형상관리

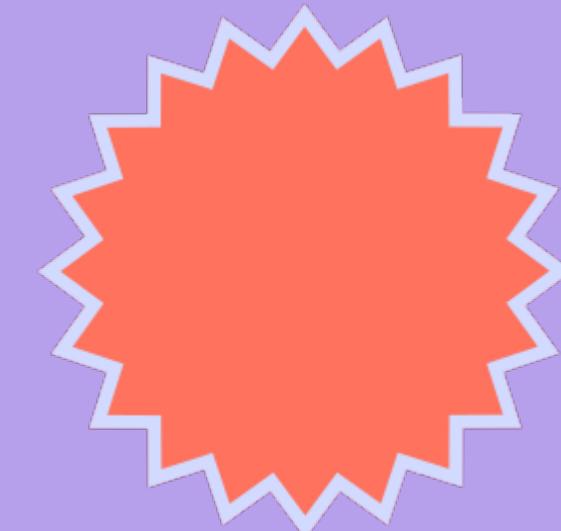
04



기능 상세



01. 원하는 여정 검색
02. 항공사 통계 및 예측 시각화 서비스
03. 리뷰 기반 분석 조회 및 리뷰 작성
04. 마이페이지



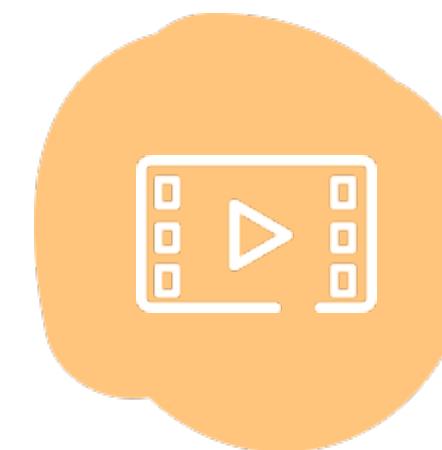
원하는 여정 검색

인천에서 출발하는 주요 도착지를 검색



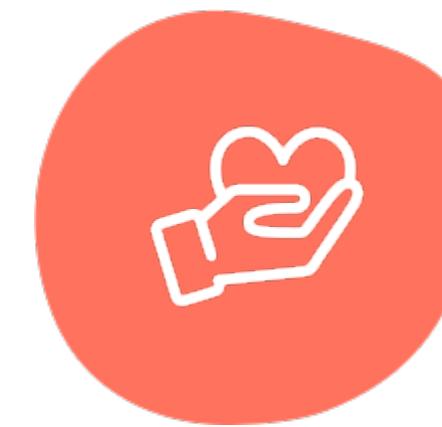
선택한 도착지 운항 항공사 리스트

선택한 도착지로 취항하는 주요 항공사의 리스트가 검색됩니다.



목적지로의 항공 데이터 미리보기

각 항공사의 운항 데이터가 미리보기 형식으로 주어집니다.
목적지로의
총 운항 횟수 / 지연률 / 평균 지연 시간 & 오늘의 예측 지연률



도시 이미지 UI

목적지를 선택하면 그에 맞는 아름다운 배경 이미지가 등장

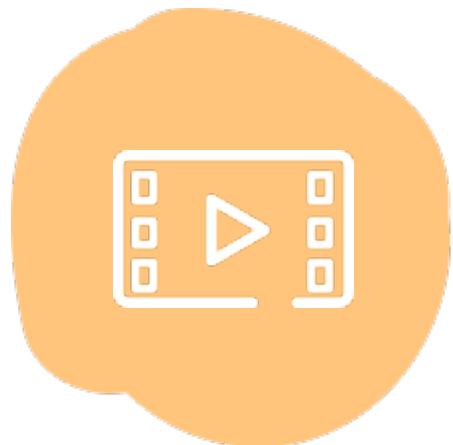
분석 리포트 Tab

선택한 항공사에 대한 데이터 분석 결과
차트로 시각화



선택한 도착지 운항 항공사 리스트

키워드 및 페이지에 대한 부연 설명을 입력해주세요.
폰트는 나눔스퀘어라운드 Regular, 크기는 24입니다.



목적지로의 항공 데이터 미리보기

키워드 및 페이지에 대한 부연 설명을 입력해주세요.
폰트는 나눔스퀘어라운드 Regular, 크기는 24입니다.



도시 이미지 UI

키워드 및 페이지에 대한 부연 설명을 입력해주세요.
폰트는 나눔스퀘어라운드 Regular, 크기는 24입니다.

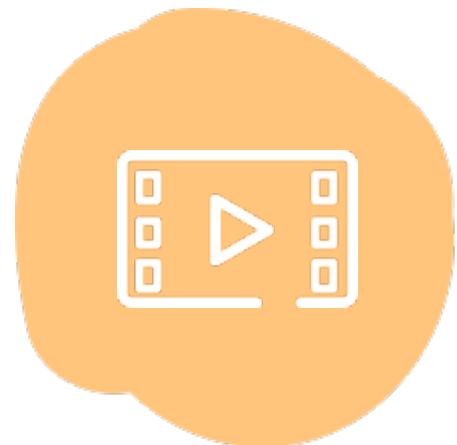
항공사 리뷰 Tab

인천에서 출발하는 주요 도착지를 검색



선택한 도착지 운항 항공사 리스트

키워드 및 페이지에 대한 부연 설명을 입력해주세요.
폰트는 나눔스퀘어라운드 Regular, 크기는 24입니다.



목적지로의 항공 데이터 미리보기

키워드 및 페이지에 대한 부연 설명을 입력해주세요.
폰트는 나눔스퀘어라운드 Regular, 크기는 24입니다.



도시 이미지 UI

키워드 및 페이지에 대한 부연 설명을 입력해주세요.
폰트는 나눔스퀘어라운드 Regular, 크기는 24입니다.

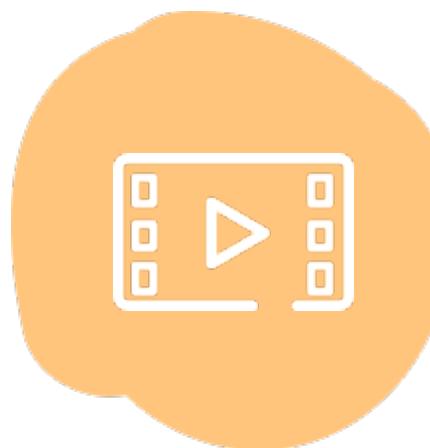
마이페이지

인천에서 출발하는 주요 도착지를 검색



선택한 도착지 운항 항공사 리스트

키워드 및 페이지에 대한 부연 설명을 입력해주세요.
폰트는 나눔스퀘어라운드 Regular, 크기는 24입니다.



목적지로의 항공 데이터 미리보기

키워드 및 페이지에 대한 부연 설명을 입력해주세요.
폰트는 나눔스퀘어라운드 Regular, 크기는 24입니다.



도시 이미지 UI

키워드 및 페이지에 대한 부연 설명을 입력해주세요.
폰트는 나눔스퀘어라운드 Regular, 크기는 24입니다.