

분류			ID	주요기능	상세기능	비고	우선순위
대분류	1단계	2단계					
데이터 등록	파일 등록	CSV	DI_CSV	원본 데이터 저장 (csv)	S3 서버에 데이터 저장	원본 데이터 저장소 향후 S3가 파일 경로를 front-end로 보내줌	
		xlsx	DI_XL	원본 데이터 저장 (xlsx)	S3 서버에 데이터 저장	-	
	분석 저장	DB	DI_RST	분석 결과 저장	열결된 DB에 분석 결과 저장	분석 결과 MySQL에 테이블 만들어서 저장	
데이터 조회 (이용자 입장)	데이터 리스트	list	DR_LI	서버에 저장된 모든 데이터 셋 게시	저장된 전체 데이터셋 조회하기		
	분석 내용 조회	detail	DR_DE	각 데이터 셋 별 분석 내용	- outlier 정제 전 후 그래프 및 분석 데이터 제시 (이용자가 선택하여 원하는 상태의 데이터 분석 제공)		
	시각화	histogram	DR_VI_HI	수치형	column의 value 별 분포		
		box plot	DR_VI_BP	수치형	q1 ~ q3		
		bar chart	DR_VI_BC	범주형	텍스트의 분포 정도 파악		
		donut chart	DR_VI_DC	true/false	y / n 데이터 비율 파악		
	컬럼 선택	select column	DR_SC	column 별 분석 내용	데이터를 보고싶은 컬럼 여려개 선택	드랍다운	
	필터링	Filter	DR_FT	데이터 범위 설정	각 컬럼별 데이터 범위 설정		
데이터 분석 (자체 분석 방법)	검색	Search	DR_SC	검색을 통해 데이터 조회	키워드 검색을 통해 특정 데이터 분석 조회하기 (Back-end에서 검색 정보 넘겨주기)	데이터 셋의 이름을 통한 검색	하
	데이터 분류	type	DA_TP	데이터 형식 파악	데이터 형식을 통해 수치형/범주형 분류	text의 경우 수치형인지 범주형인지 어떻게 판단할 수 있을지?	
	수치형 feature 추출	Basic Statistics	DA_NM_BS	기초 통계치 파악	평균, 중앙값, 최빈값, 표준편차, 분산, 최소/최대, vif	pySpark 에서드 사용함 (pandas describe)	
		Data Distribution	DA_NM_RG	데이터 분포 파악	q1, q2, q3 값 및 q1 ~ q3 범위 내 데이터 파악		
	범주형 feature 추출	Chi Sqared Test	DA_CT_CHI	카이제곱 검정	적합성 / 독립성 검정	- 하나의 범주형 변수에 대해 관측값들이 어떤 분포를 띄는지 검정 - 서로 다른 두 범주형 변수 간에 연관성이 있는지 검정	
	이상치 처리	Outlier	DA_OL	이상치 처리 기준	- outlier 정제 기준: - null 값 정제 기준:	논문, 기술 서적 학습 후 기준 정립 예정	
부가 기능	사이트 기본 틀	footer	SB_FT	사이트 하단 footer	all rights by 000	굳이 만들어야 할지? ..푸터는 사랑입니다 저희 팀을 봐야해야 해요	중
		nav	SB_NV	사이트 상단 nav	전체 리스트 조회, 검색, 데이터 추가 기능		
	14	20					