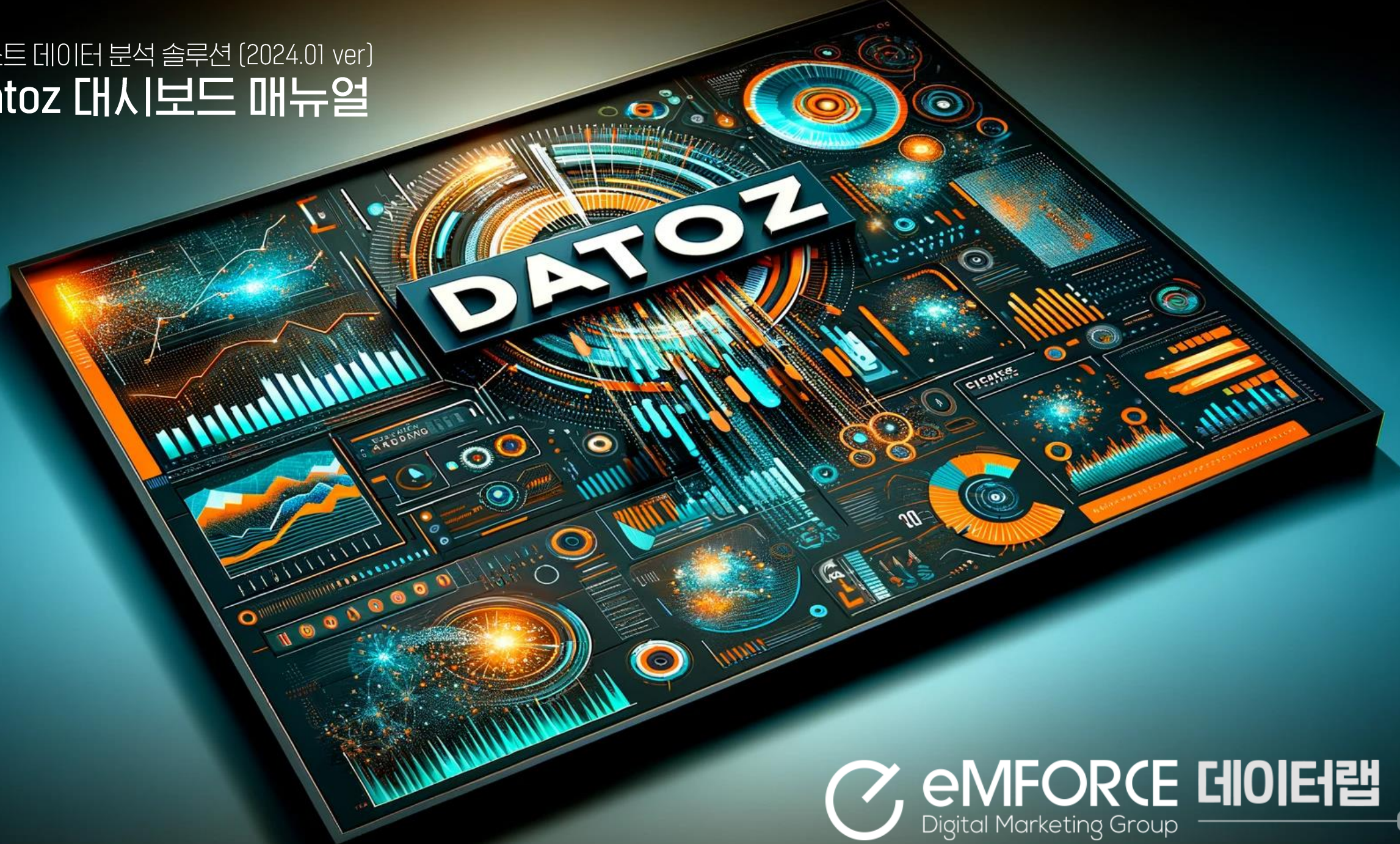


텍스트 데이터 분석 솔루션 (2024.01 ver)
Datoz 대시보드 매뉴얼



데이터의 인사이트를 알려줄게! 데이토즈



솔루션 접속 계정 : <http://43.200.206.201:3838/Datoz/>

데이터 기반의 실천적 마케팅을 추구하는 엠포스가
2024년 새해를 맞아 또 하나의 새로운 솔루션을 **런칭**했습니다.

사실 엠포스는 이미 지난 2022년 8월,
Datoz(데이토즈)라는 동일한 이름의 분석 솔루션을 선보인 바 있는데요.

별도의 광고나 홍보를 하지 않았는데도 불구하고 많은 분들이 자발적으로 찾고
또 꾸준히 이용해 주신 덕분에 새롭게 업그레이드 된 버전을 내놓게 되었습니다.

<데이토즈 솔루션의 특징점>

- 별도의 **프로그램 설치나 회원가입 필요 없이** 공유된 URL로 접속하여 이용
- 상품 후기, 홈페이지 게시물, 댓글 등 개인 및 기업이 보유한, 혹은 크롤링한 다양한 **텍스트(Text) 데이터**를 업로드해서 분석 가능
- **코딩에 대한 지식 없이도** 클릭 몇 번으로 **고급 통계 분석**을 자유롭게~
- 분석 기법 : Text Mining, WordCloud, SNA, TF-IDF, LDA, N-gram 등
- 엑셀 포맷의 텍스트 데이터라면 **용량이나 양식에 상관없이** 이용 가능
- 결정적으로, 솔루션의 **모든 기능은 100% 무료**로 이용할 수 있습니다!!

열거한 특징들로 보시는 것처럼, 이번 버전에서는

기존 버전에서 다소 **불편하다고 말씀 주셨던 많은 부분을 개선**하는데 힘썼고
기존 버전에 없던 **새로운 분석 기법들을 대폭 추가해 업그레이드** 하였습니다.

엠포스는 이 외에도

- 다양한 형태의 데이터를 자유롭게 분석할 수 있는 솔루션을 개발 중이며,
- 데이토즈를 조금씩, 하지만 꾸준히 업그레이드 해 나가는 것은 물론,
 - 온라인광고 운영 데이터 등을 기반으로 한 분석 솔루션도 곧 런칭 예정입니다.

엠포스가 개발하는 솔루션들은 모두
기술이 아닌 **실제 데이터 분석가의 노하우를 모델로** 하고 있습니다.

데이터 분석이 필요한 모든 분들이 마음껏 이용해주시고
또 마음껏 리뷰를 들려주시면 언젠가, “코딩 따위 몰라도 데이터 분석이 가능한!” 세상이 오겠지요.



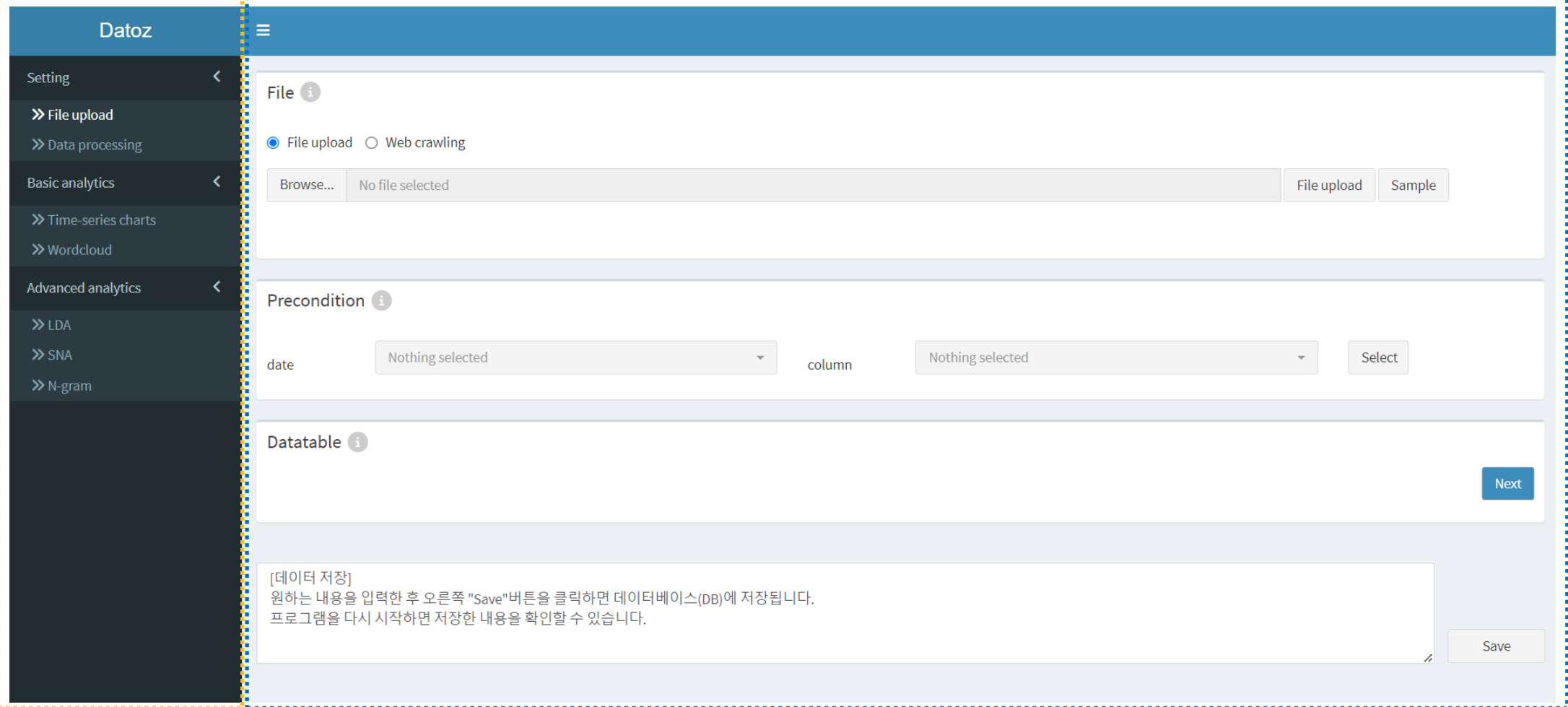
...2024.01.03 엠포스 데이터전략실에서

Datoz 대시보드 매뉴얼

기본 화면 구성

목차

메인화면



Datoz

Setting <

- >> File upload
- >> Data processing

Basic analytics <

- >> Time-series charts
- >> Wordcloud

Advanced analytics <

- >> LDA
- >> SNA
- >> N-gram

File ⓘ

File upload Web crawling

Browse... No file selected File upload Sample

Precondition ⓘ

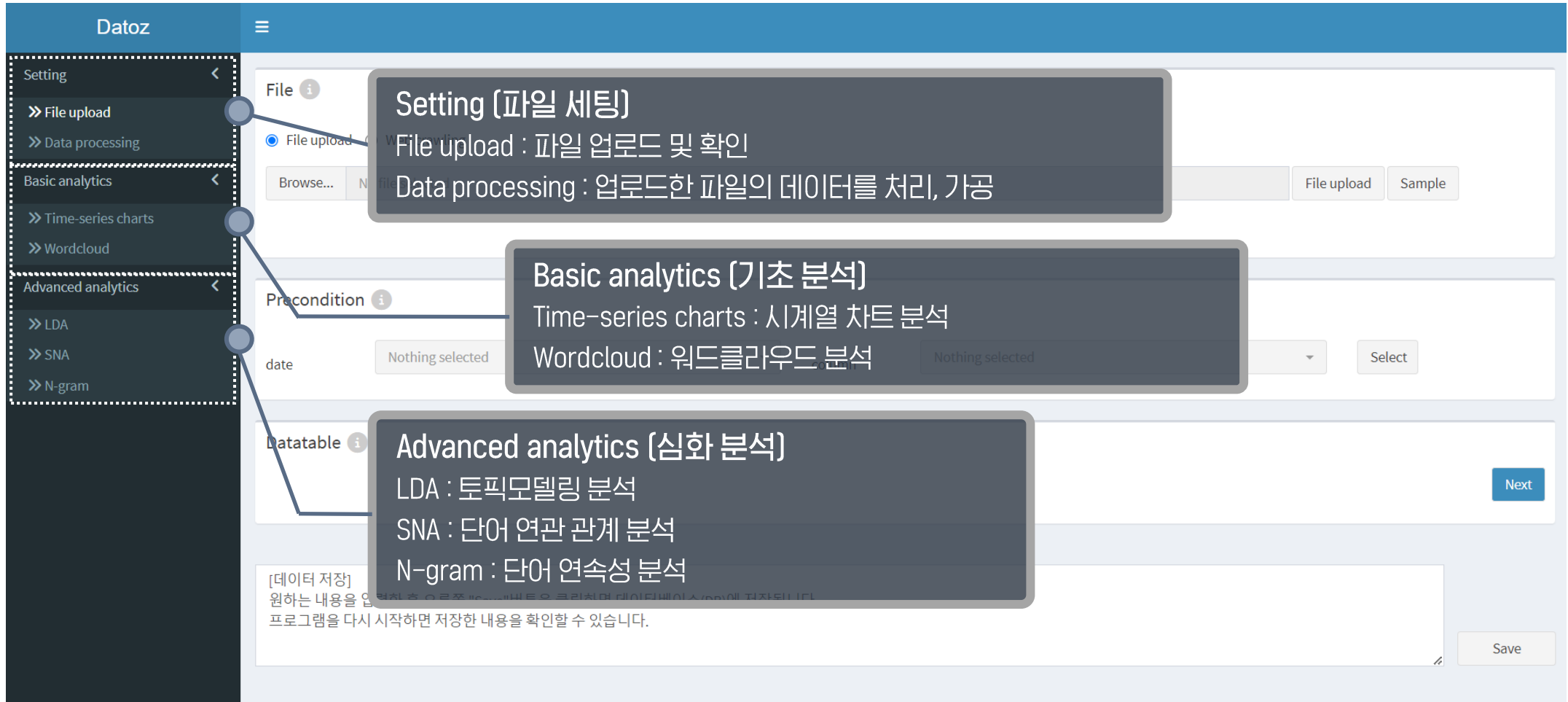
date Nothing selected column Nothing selected Select

Datatable ⓘ

Next

[데이터 저장]
원하는 내용을 입력한 후 오른쪽 "Save" 버튼을 클릭하면 데이터베이스(DB)에 저장됩니다.
프로그램을 다시 시작하면 저장한 내용을 확인할 수 있습니다.

Save

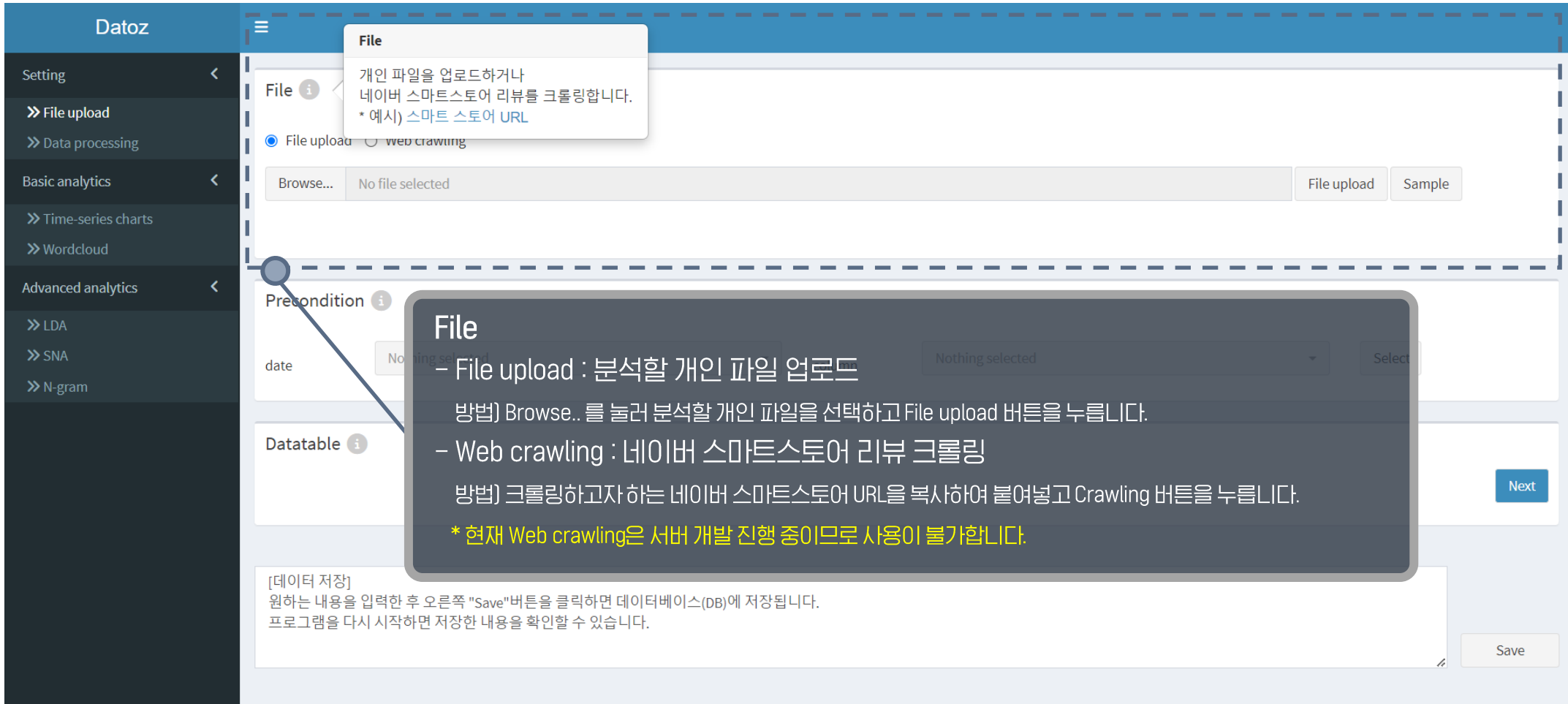


The screenshot shows the Datoz dashboard interface. On the left is a dark sidebar menu with categories: Setting, Basic analytics, and Advanced analytics. Three callout boxes point to specific items in the menu:

- Setting (파일 세팅)**
 - File upload : 파일 업로드 및 확인
 - Data processing : 업로드한 파일의 데이터를 처리, 가공
- Basic analytics (기초 분석)**
 - Time-series charts : 시계열 차트 분석
 - Wordcloud : 워드클라우드 분석
- Advanced analytics (심화 분석)**
 - LDA : 토픽모델링 분석
 - SNA : 단어 연관 관계 분석
 - N-gram : 단어 연속성 분석

The main content area shows a 'File' section with 'File upload' selected, a 'Precondition' section with a 'date' field, and a 'Datatable' section. A 'Next' button is visible at the bottom right of the main content area, and a 'Save' button is at the bottom right of the entire page.

Setting >> File upload



The screenshot shows the 'File upload' configuration page in the Datoz dashboard. The left sidebar contains navigation options: Setting, File upload, Data processing, Basic analytics, and Advanced analytics. The main content area has two radio buttons: 'File upload' (selected) and 'web crawling'. Below these are 'Browse...' and 'No file selected' buttons, and 'File upload' and 'Sample' buttons. A 'Next' button is at the bottom right. A 'Save' button is at the bottom right of a 'Precondition' section. A 'Datatable' section is also visible.

File
개인 파일을 업로드하거나 네이버 스마트스토어 리뷰를 크롤링합니다.
* 예시) 스마트 스토어 URL

File
- File upload : 분석할 개인 파일 업로드 방법) Browse.. 를 눌러 분석할 개인 파일을 선택하고 File upload 버튼을 누릅니다.
- Web crawling : 네이버 스마트스토어 리뷰 크롤링 방법) 크롤링하고자 하는 네이버 스마트스토어 URL을 복사하여 붙여넣고 Crawling 버튼을 누릅니다.
*** 현재 Web crawling은 서버 개발 진행 중이므로 사용이 불가합니다.**

[데이터 저장]
원하는 내용을 입력한 후 오른쪽 "Save" 버튼을 클릭하면 데이터베이스(DB)에 저장됩니다.
프로그램을 다시 시작하면 저장한 내용을 확인할 수 있습니다.

Datoz 대시보드 매뉴얼

Setting >> File upload

Datoz

- Setting <
- >> File upload
- >> Data processing
- Basic analytics <
- >> Time-series charts
- >> Wordcloud
- Advanced analytics <
- >> LDA
- >> SNA
- >> N-gram

File ⓘ

File upload Web crawling

Browse... No file selected File upload Sample

Precondition

Sample 버튼을 누르면,
데이터 형식을 확인할 수 있는 샘플자료가 보여집니다.

Sample

사용 가능한 파일 형식 : **xlsx, csv**
해당 데이터는 커뮤니티 글의 일부를 발췌한 샘플자료입니다. (100자 이내)

id	날짜	제목	본문
0	2021.01.04	우리집이 돼지갈비 맛집(0)	얼마전 하나로마트에서 목살을 세일했어요 목살을 세일하면 뭐해야한다 돼지갈비 재워야조 최근에 꽃힌건 주야공님 돼지갈비라서 엄청 자주 요리했었어요 둘째아이가 어려서인지 고기 두께가 굵
1	2021.01.04	주방용품 생활용품 집화 500원부터~	택배거래시 착불 또는 선불 원 예민함 교환 환불 예누리 사절 새제품은 명시되어있습니다 개인 위생용품 새제품 가지 일괄 원 손살균보습제 년 월까지 순세정월 년 월까지 닥터클린 스프레이
2	2021.01.04	고양이가 습식 먹고 눈물 흘렸어요 ㅋㅋ	개월 캣초딩이에요 밥 먹으면서 타액 분비 신경이랑 눈물 분비 신경이 혼선되서 눈물 흘릴 수도 있다면서요 습식 먹다 눈물 흘린다고 하는 말만 들어봤지 전엔 못봤는데 오늘 습식 먹이고
3	2021.01.04	락앤락 후라이팬에 콩나물 불고기^^	주부는 아니지만 저녁거리는 늘 고민이죠 오늘 저녁에 먹은 건 아니지만 메뉴 고민일때 백종원 유튜브 요리비책 보면서 요리하기 쉽고 재료도 쉽게 접할 수 있는 거면 한 번씩 도전해요
4	2021.01.03	후르츠 칵테일 스케일	남편이 후르츠 칵테일 먹고싶대서 쿠팡으로 시켰어요 택배를 열어보고 그 크기에 감놀 업소용 인가봐요 얼마난건줄 알고 산거나니깐 그런건 안보고 그냥 샀대요 뭘 살때 가격을 안봐도 그냥

Showing 1 to 5 of 5 entries Previous 1 Next

Precondition

Sample 버튼을 누르면,
데이터 형식을 확인할 수 있는 샘플자료가 보여집니다.

Sample

사용 가능한 파일 형식 : **xlsx, csv**
* **xlsx** 는 첫 시트만 사용합니다.

필요한 데이터

- 1) 시계열 분석에 사용할 날짜 변수
- 2) 기초/심화 분석에 사용할 텍스트로 이루어진 변수
ex) 제목 or 본문

* 날짜와 텍스트 컬럼을 제외한 다른 컬럼은 없어도 무방

Datoz 대시보드 매뉴얼

Setting >> File upload

Datoz

- Setting <
- >> File upload
- >> Data processing
- Basic analytics <
- >> Time-series charts
- >> Wordcloud
- Advanced analytics <
- >> LDA
- >> SNA
- >> N-gram

File ⓘ

File upload Web crawling

Browse... sample.csv File upload Sample

Upload complete

Precondition ⓘ

date Nothing selected Select

Datatable ⓘ

No.	V1	id.1	사이트명	날짜	제목	본문	사용자	Remove
						얼마전 하나로마트에서 특별 할인 행사를 해서... (복합제거 관련 설명)		

File upload를 클릭하면,
아래의 Datatable에서
업로드한 파일을 확인할 수 있습니다.

- * 업로드한 파일에 중복 데이터가 있는 경우,
중복제거로 인해 데이터 건수가 달라질 수 있습니다.
- * 중복제거를 원치 않을 경우,
key값이 있는 데이터를 업로드해주세요.

Datoz 대시보드 매뉴얼

Setting >> File upload

File

File upload Web crawling

Browse... sample.csv File upload Sample

Upload complete

Precondition

date 날짜 column 제목, 본문 Select

V1 id.1 사이트명 날짜 제목 본문 사용자

Next Remove 사용자

Precondition

- date : 날짜에 해당하는 변수명 선택
날짜 변수가 없을 경우, 'None' 선택
** None을 선택한 경우, 시계열 차트 분석 불가*
- column : 텍스트분석을 진행할 변수명 선택
** column은 다중선택 가능*

Datoz 대시보드 매뉴얼

Setting >> File upload

Advanced analytics <

- >> LDA
- >> SNA
- >> N-gram

Precondition ⓘ

date: 날짜 column: 제목, 본문 Select

Include: Connect words with comma


Exclude: Connect words with comma

Apply

Datatable ⓘ

No.	날짜	제목	본문	Remove
1	2021.01.04	우리집이 돼지갈비 맛집(0)	<p>얼마전 하나로마트에서 목살을 세일했어요 목살을 세... 자주 요리했었어요 ㅎㅎ 둘째야기가 어려서인지 고기 두... 네요 ㅎㅎ 그러다가 한주부님 돼지갈비도 그리워 이번엔... 데 몇번하고나니 어렵지도 않아요 ㅎㅎ 손에 익은듯 착... 간장만 1t스푼쯤 간보며 먹은듯합니다 그러다보면 나중에는 찐지?단지? 감각이 없어지지요 ㅋㅋ 계량 요리의 장점입니다 ㅎㅎ 락앤락에서 얼마전 퍼플에디션 세일을해서 1.8리터를 9900원에 세일하길래 2개 들어봤는데 고기 1키로가 딱! 들어갑니다 ㅎㅎ 오랜만에 먹는 한주부님 갈비는 사랑입니다 ㅎㅎ 시레갈비는 다들 너무나 맛있어서 늘 누구레시피를 할지 즐거운 고민에 빠집니다. 레시피 제공해주신 모든분들 감사하고 사랑합니다 ~^^♥</p> <p>택배거래시 착불 또는 선불 3,500원 예민망 교환 환불 예누리 사절 새제품은 명시되어있습니다 1.개인위생용품 새제품 3가지 일괄 10,000원 -손 살균보습제 70ml 2023년 2월까지 -손세정젤 60ml 2023년 3월까지 -탁티클린 스프레이 100ml 2023년 3월까지 2.어린이들 발재기자 새제품 1,000원 3.스카트 일회용행주 5롤(두께4) 5,000원 - 판매완료 4.투명위생컵 200ml 새제품 5개 500원 5.열탕집게 1,000원 6.보온보냉주머니 새제품 1,000원 7.마스크스트랩 새제품 1,000원 8.칼도마정리대 11cm*11cm*20cm 새제품 2,000원 9.헝겝 칼블럭 5,000원(약간의 기름됨 있습니다) 10.실리 쿡 유리용기 새제품 3개 일괄 9,000원 -제일우측 물병 약 1리터 들어갑니다 12.전자렌지 스티미세척기 앵그리마미 새제품 2,000원 13.JMW 헤어드라이어 거치대 사이즈 76mm*84mm*67mm 새제품 3,000원 14.포장용품 일괄 10,000원 왼쪽-손잡이박스2개 끈 포함 15*15*24 오른쪽-넓적한박스 2개 25*20*8 마끈,스티커,민트색포장끈2개,크라프트스타핑2개 15.핸드폰거치대 새제품 3,000원 16.쿠쿠 찜빔침대 20.5cm*5cm 새제품 2,000원 17.</p>	Remove

Select를 클릭하면,
아래의 Datatable에서
선택한 날짜와 컬럼만을 추출한
데이터 테이블을 확인할 수 있습니다.



Datoz 대시보드 매뉴얼

Setting >> File upload

Advanced analytics <
Precondition i

date

Include

Exclude

날짜

레시피, 요리

택배&착불&환불&새제품

column

제목, 본문

Select

Apply

Precondition

- Include : 내용에 포함되어야 할 단어 기입
- Exclude: 내용에 포함되지 않아야 할 단어 기입

Ex) '레시피' 또는 '요리'가 내용에 포함되고, '택배', '착불', '환불', '새제품'이라는 단어가 모두 포함된 경우 행 제거.

* 콤마(,)로 연결 시 '또는' 조건
 앤드(&)로 연결 시 '그리고' 조건이지만,
 한 칸에 두 가지를 혼용하여 사용할 수 없음.

	제목	본문	
4	2021.01.04	<p>얼마전 하나로마트에서 목살을 세일했어요 목살을 세일하면 뭐해야한다? 돼지갈비 재워야죠? ㅎㅎ 최근에 꽃힌건 주아공님 돼지갈비라서 엄청 자주 요리했었어요 ㅎㅎ 둘째아기가 어려서인지 고기 두께가 굵은건 잘라줘도 잘 씹지 못하기에 자주 대패삼겹살 버전으로 앞다리 돼지갈비를 했었는데 ㅎㅎ 그러다가 한주부님 돼지갈비도 그리워 이번엔 목살세일 기념으로 갈비를 재워줬어요~ 처음 했을때는 뭔가 과정이 복잡하듯 느껴졌었는데 몇 번하고나니 어렵지도 않아요 ㅎㅎ 손에 익은듯 착착~! 잘 해냅니다 ㅎㅎ 계량으로 요리를 하니 간을 보지않는게 좋아요~ 예전에는 갈비하나 하면 간장만 1t스푼쯤 간보며 먹은듯합니다 그러다보면 나중에는 찐지?단지? 감감이 없어지지요 ㅋㅋ 계량 요리의 장점입니다 ㅎㅎ 락앤락에서 얼마전 퍼플에디션 세일을해서 1.8리터를 9900원에 세일하길래 2개 들여놨는데 고기 1키로가 딱! 들어갑니다 ㅎㅎ 오랜만에 먹는 한주부님 갈비는 사랑입니다 ㅎㅎ 시레갈비는 다들 너무나 맛있어서 늘 누구레시피를 할지 즐거운 고민에 빠집니다. 레시피 제공해주신 모든분들 감사하고 사랑합니다~~^^♥</p>	Remove
		<p>주부는 아니지만 저녁거리는 늘 고민이죠 ㅎㅎ 오늘 저녁에 먹은 건 아니지만 메뉴 고민일때 백종원 유튜브 요리비책 보면서 요리하기 쉽고 재료도 쉽게 접할 수 있는 거면 한 번씩 도전해요^^ 지난주에 콩나물 불고기 만들어봤어요 ㅎㅎ 식구끼리 먹지만 휴대용 버너 위에다가 세팅해서 조리하면서 먹었어요^^ 식구들 반응이 좋아서 같이 밥 먹을때 다음에 또 만들어서 먹으려구요 ㅎㅎ 팬은 락앤락 제품이에요 ㅎㅎ 하드앤라이트 후라이팬 30cm 제품인데 평일에는 둘이서 먹지만 주말에는 네식구 모이면 팬이 큰 거 필요해서 작년 봄에 구매한 건데 잘 쓰고 있어요^^ 콩불이 숨이 죽으면 양이 팍 줄어들지만 처음에 콩나물 깔고 냉동 대패 올리려면 넉넉한 사이즈 팬이 필요한데 이럴 때 딱이더라구요^^ 콩불은 밖에서 사먹었는데 집에서 만들어 먹으니 좋더라구요 ㅎㅎ 부모님도 만족하시고 마지막에 밥까지 볶아 먹으니 든든하더라구요^^ 레시피로 양념장 만드면서 간 보고 좀 더 추가하고 빨간 빼고 채소도 집에 양송이 버섯 있어서 몇개 넣고 고기 먹을때 파채 같이 먹으면 좋을 거 같아서 대파도 등성등성 잘라서 넣지만 파채로 위에 뿌리고 중간중간 파채 넣어서 먹음 아삭하고 대패삼겹 느끼함도 잡아줘서 좋더라구요^^ 저녁거리 고민일 때 백종원 유튜브 보면서 입에 맞</p>	Remove

Datoz 대시보드 매뉴얼

Setting >> File upload

Advanced analytics <

- >> LDA
- >> SNA
- >> N-gram

Precondition ⓘ

date: 날짜 column: 제목, 본문 Select

Include: 레시피, 요리

Exclude: 택배&착불&환불&새제품 Apply

Datatable ⓘ

Datatable
업로드한 파일 또는
데이터 추출 및 가공 처리한
데이터 테이블을 확인할 수 있습니다.
*** Remove 버튼 클릭시, 해당 행 제거 가능**

No.	날짜	제목	본문	Remove
4	2021.01.04	락앤락 후라 이팬에 콩나물 불고기^^	<p>일마전 하나로마트에서 목살을 세일했어요 목살을 세일하면 뭐해야한다? 돼지갈비 재워야죠? ㅎㅎ 최근에 꽃힌건 주아공님 돼지갈비라서 참 좋아했어요 ㅎㅎ 둘째아기 어려서인지 고기 두께가 굵은건 잘라줘도 잘 씹지 못하기에 자주 대패삼겹살 버전으로 앞다리 돼지갈비를 했었는데 갈비도 그리워 이번엔 목살세일 기념으로 갈비를 재워줬어요~ 처음 했을때는 뭔가 과정이 복잡하듯 느껴졌었는데 몇 손에 익은듯 착착~! 잘 해냅니다 ㅎㅎ 계량으로 요리를 하니 간을 보지않는게 좋아요~예전에는 갈비하나 하면 간장 다 그러다보면 나중에는 짠지?단지? 감각이 없어지지요 ㅋㅋ 계량 요리의 장점입니다 ㅎㅎ 락앤락에서 얼마전 퍼플에 10원에 세일하길래 2개 들여놨는데 고기 1키로가 딱! 들어갑니다 ㅎㅎ 오랜만에 먹는 한주부님 갈비는 사랑입니다 ㅎㅎ ㅎㅎ 시레갈비는 다들 너무나 맛있어서 늘 누구레시피를 할지 즐거운 고민에 빠집니다. 레시피 제공해주신 모든분들 감사하고 사랑합니다~~^^♥</p> <p>주부는 아니지만 저녁거리는 늘 고민이죠 ㅎㅎ 오늘 저녁에 먹은 건 아니지만 메뉴 고민일때 백종원 유튜브 요리비책 보면서 요리하기 쉽고 재료도 쉽게 접할 수 있는 거면 한 번씩 도전해요^^ 지난주에 콩나물 불고기 만들어봤어요 ㅎㅎ 식구끼리 먹지만 휴대용 버너 위에다가 세팅해서 조리하면서 먹었어요^^ 식구들 반응이 좋아서 같이 밥 먹을때 다음에 또 만들어서 먹으려구요 ㅎㅎ 팬은 락앤락 제품이에요 ㅎㅎ 하드앤라이트 후라이팬 30cm 제품인데 평일에는 둘이서 먹지만 주말에는 네식구 모이면 팬이 큰 거 필요해서 작년 봄에 구매한 건데 잘 쓰고 있어요^^ 콩불이 숨이 죽으면 양이 팍 줄어들지만 처음에 콩나물 깔고 냉동 대패 올리려면 넉넉한 사이즈 팬이 필요한데 이럴 때 딱이더라구요^^ 콩불은 밖에서 사먹었는데 집에서 만들어 먹으니 좋더라구요 ㅎㅎ 부모님도 만족하시고 마지막에 밥까지 볶아 먹으니 든든하더라구요^^ 레시피로 양념장 만드면서 간 보고 좀 더 추가하고 빨간 빼고 채소도 집에 양송이 버섯 있어서 몇개 넣고 고기 먹을때 파채 같이 먹으면 좋을 거 같아서 대파도 등성등성 잘라서 넣지만 파채로 위에 뿌리고 중간중간 파채 넣어서 먹음 아삭하고 대패삼겹 느끼함도 잡아줘서 좋더라구요^^ 저녁거리 고민일 때 백종원 유튜브 보면서 입에 맞</p>	<div style="text-align: right;"> Remove </div>

Datoz 대시보드 매뉴얼

Setting >> File upload

10	2021.01.03	\uD83D\uDC1B 그 때 얘기한 회 사ㄷㄹㅇ 뉴설 \uD83D\uDE28	<p>이제부터는... (text continues)</p>	Remove
----	------------	--	----------------------------------	--------

Save를 클릭하면,
해당 내용이 저장되고 프로그램을 다시 시작할 경우,
저장 내용을 확인 할 수 있는 메모장 기능

*** 현재 데이터 저장기능은 서버 개발 진행 중이므로 사용이 불가합니다.**

[데이터 저장]
원하는 내용을 입력한 후 오른쪽 "Save" 버튼을 클릭하면 데이터베이스(DB)에 저장됩니다.
프로그램을 다시 시작하면 저장한 내용을 확인할 수 있습니다.

Save

Datoz 대시보드 매뉴얼

Setting >> File upload

Advanced analytics <

- >> LDA
- >> SNA
- >> N-gram

Precondition ⓘ

date: 날짜 column: 제목, 본문

Select

Include: 레시피, 요리

Exclude: 택배&착불&환불&새제품

Apply

Datatable ⓘ

No.	날짜	제목	본문	Remove
1	2021.01.04	우리집이 돼지갈비 맛집(O)	얼마전 하나로마트에서 목살을 세일했어요 목 주 요리했었어요 ㅎㅎ 둘째아기가 어려서인지 고요 ㅎㅎ 그러다가 한주부님 돼지갈비도 그리워이 변하고나니 어렵지도 않아요 ㅎㅎ 손에 익은듯 ㅎㅎ 잘 먹었어요 ㅎㅎ 요리를 할 때 정말 맛있는 거 같아요 ㅎㅎ 저는 갈비 하나만 먹으면 간장만 1t스푼쯤 간보며 먹은듯합니다 그러다보면 나중에는 찐지?단지? 감각이 없어지지요 ㅋㅋ 계량 요리의 장점입니다 ㅎㅎ 락앤락에서 얼마전 퍼플에디션 세일을해서 1.8리터를 9900원에 세일하길래 2개 들여놨는데 고기 1키로가 딱! 들어갑니다 ㅎㅎ 오랜만에 먹는 한주부님 갈비는 사랑입니다 ㅎㅎ 시레갈비는 다들 너무나 맛있어서 늘 누구레시피를 할지 즐거운 고민에 빠집니다. 레시피 제공해주신 모든분들 감사하고 사랑합니다~~^^♥	Remove
4	2021.01.04	락앤락 후라이팬에 콩나물 불고기^^	주부는 아니지만 저녁거리는 늘 고민이죠 ㅎㅎ 오늘 저녁에 먹은 건 아니지만 메뉴 고민일때 백종원 유튜브 요리비책 보면서 요리하기 쉽고 재료도 쉽게 접할 수 있는 거면 한 번씩 도전해요^^ 지난주에 콩나물 불고기 만들어봤어요 ㅎㅎ 식구끼리 먹지만 휴대용 버너 위에다가 세팅해서 조리하면서 먹었어요^^ 식구들 반응이 좋아서 같이 밥 먹을때 다음에 또 만들어서 먹으려구요 ㅎㅎ 팬은 락앤락 제품이에요 ㅎㅎ 하드앤라이트 후라이팬 30cm 제품인데 평일에는 둘이서 먹지만 주말에는 네식구 모이면 팬이 큰 거 필요해서 작년 봄에 구매한 건데 잘 쓰고 있어요^^ 콩불이 숨이 죽으면 양이 팍 줄어들지만 처음에 콩나물 깔고 냉동 대패 올리려면 넉넉한 사이즈 팬이 필요한데 이럴 때 딱이더라구요^^ 콩불은 밖에서 사먹었는데 집에서 만들어 먹으니 좋더라구요 ㅎㅎ 부모님도 만족하시고 마지막에 밥까지 볶아 먹으니 든든하더라구요^^ 레시피로 양념장 만드면서 간 보고 좀 더 추가하고 빨간 빼고 채소도 집에 양송이 버섯 있어서 몇개 넣고 고기 먹을때 파채 같이 먹으면 좋을 거 같아서 대파도 등성등성 잘라서 넣지만 파채 루에 뿌리고 중간중간 파채 넣어서 먹음 아삭하고 대패삼겹 느끼함도 잡아줘서 좋더라구요^^ 저녁거리 고민일 때 백종원 유튜브 보면서 입에 맞	Remove

Next를 누르면,
 Setting(파일 세팅)의
 Data processing 목차로 넘어갑니다.

Setting >> Data processing

Tokenization
 기존 선택한 텍스트 컬럼의 토큰화 과정을 확인할 수 있는 데이터 테이블.
 - reviews 컬럼 : 기존 선택한 텍스트 컬럼의 결합
 - token 컬럼 : 명사 단위로 쪼갬 토큰들

Token 컬럼의 데이터를 더블 클릭하면, 직접 내용을 수정하여 원하는 단어를 추가하거나 삭제할 수 있습니다.
 * 토큰은 항상 콤마(,)로 연결

Setting >> Data processing

The screenshot shows the Datoz dashboard interface. On the left is a dark sidebar menu with the following items: Setting, File upload, Data processing (highlighted), Basic analytics, Time-series charts, Wordcloud, Advanced analytics, LDA, SNA, and N-gram. The main content area is titled 'Dictionary edit' and contains two input fields for 'Add' and 'Delete', both containing the text '개발중입니다.' (Under development), and an 'Apply' button. Below this is a 'Tokenization' section with a table header 'id reviews'. The first row of the table contains a paragraph of Korean text. A callout box with a blue arrow pointing to a 'Next' button in the bottom right corner contains the text: 'Next를 누르면, Basic analytics(기초 분석)의 Time-series charts 목차로 넘어갑니다.' (When you click Next, you will go to the Time-series charts table of contents in Basic analytics (Basic Analysis)).

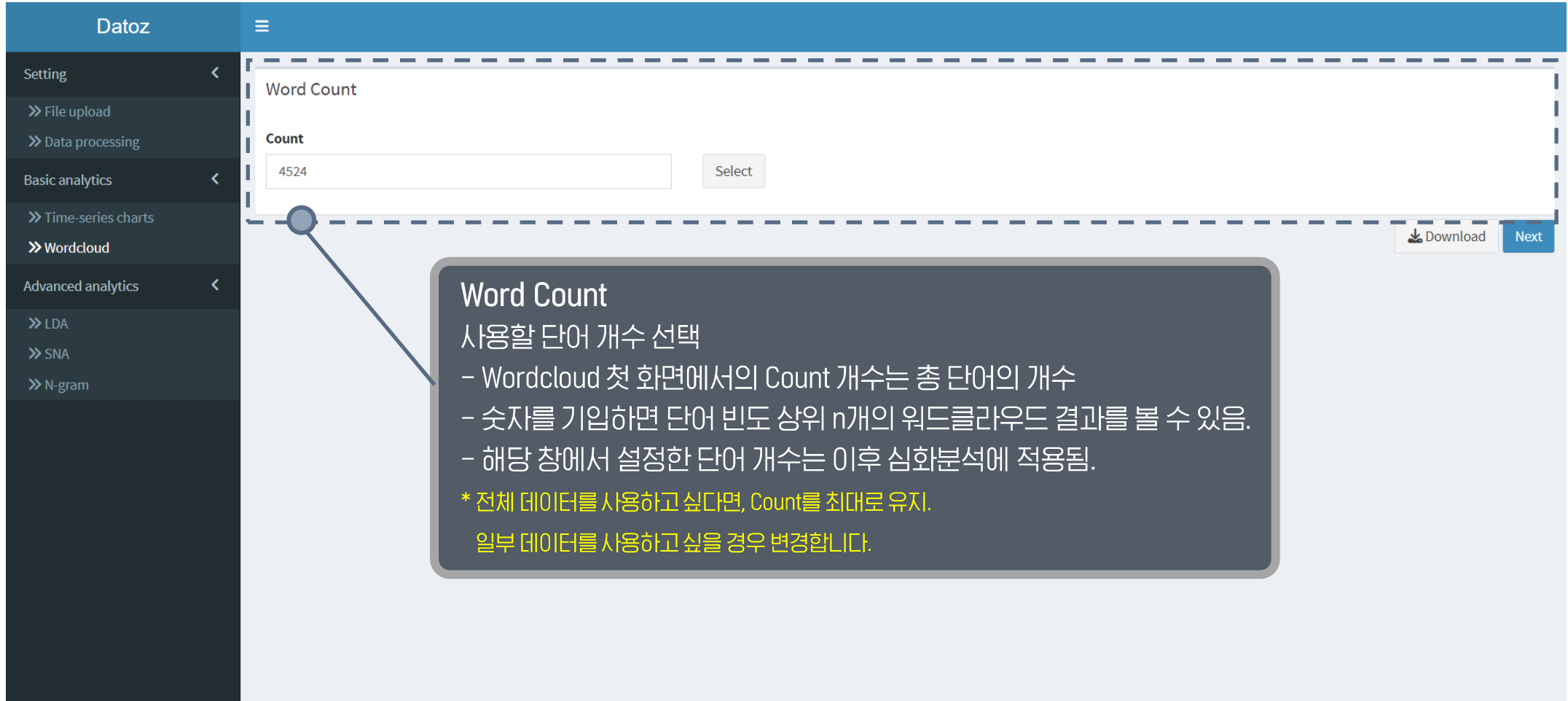
Basic analytics >> Time-series charts

The screenshot shows the Datoz dashboard interface. On the left is a navigation menu with categories: Setting, Basic analytics, and Advanced analytics. The 'Basic analytics' section is expanded to show 'Time-series charts' and 'Wordcloud'. The main content area displays a 'Time-series chart' with a 'date category' dropdown menu set to 'day'. The chart shows a fluctuating blue line representing frequency over time. A 'Next' button is highlighted with a dashed box. Two callout boxes provide instructions: one above the 'Next' button and one below the chart.

Next를 누르면,
Wordcloud 목차로 넘어갑니다.

시계열 추이는 day, month, year 구분으로
각각 일별, 월별, 년도별 데이터 건수의 추이를
확인할 수 있습니다.
* 날짜변수가 없는 경우, 시계열 차트를 볼 수 없습니다.

Basic analytics >> Wordcloud



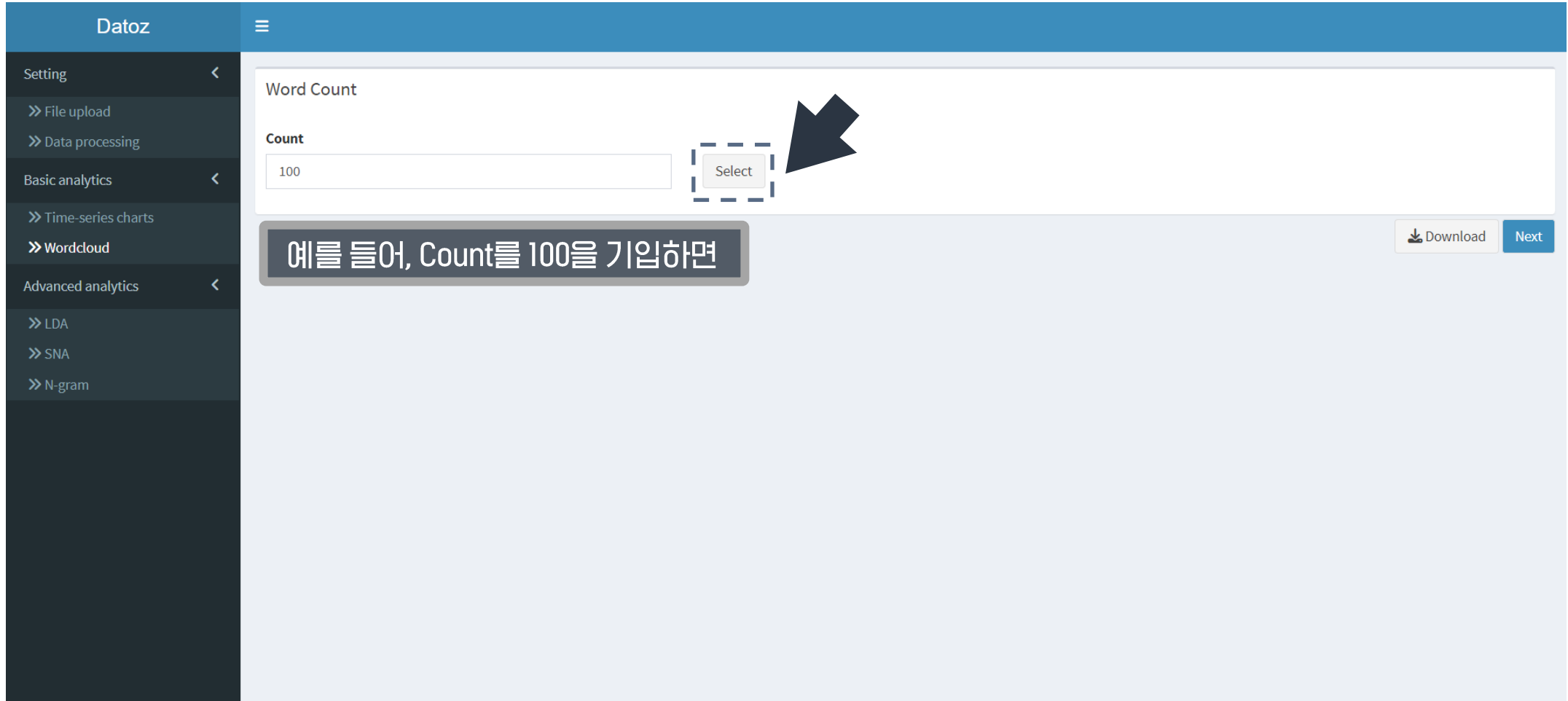
The screenshot shows the Datoz dashboard interface. On the left is a dark sidebar with navigation options: Setting, Basic analytics, and Advanced analytics. The main content area is titled 'Word Count' and features a 'Count' input field with the value '4524' and a 'Select' button. A 'Next' button is located at the bottom right of the main content area. A callout box points to the 'Count' field with the following text:

Word Count
사용할 단어 개수 선택

- Wordcloud 첫 화면에서의 Count 개수는 총 단어의 개수
- 숫자를 기입하면 단어 빈도 상위 n개의 워드클라우드 결과를 볼 수 있음.
- 해당 창에서 설정한 단어 개수는 이후 심화분석에 적용됨.

* 전체 데이터를 사용하고 싶다면, Count를 최대로 유지.
일부 데이터를 사용하고 싶을 경우 변경합니다.

Basic analytics >> Wordcloud



Word Count

Count

100

Select

예를 들어, Count를 100을 기입하면

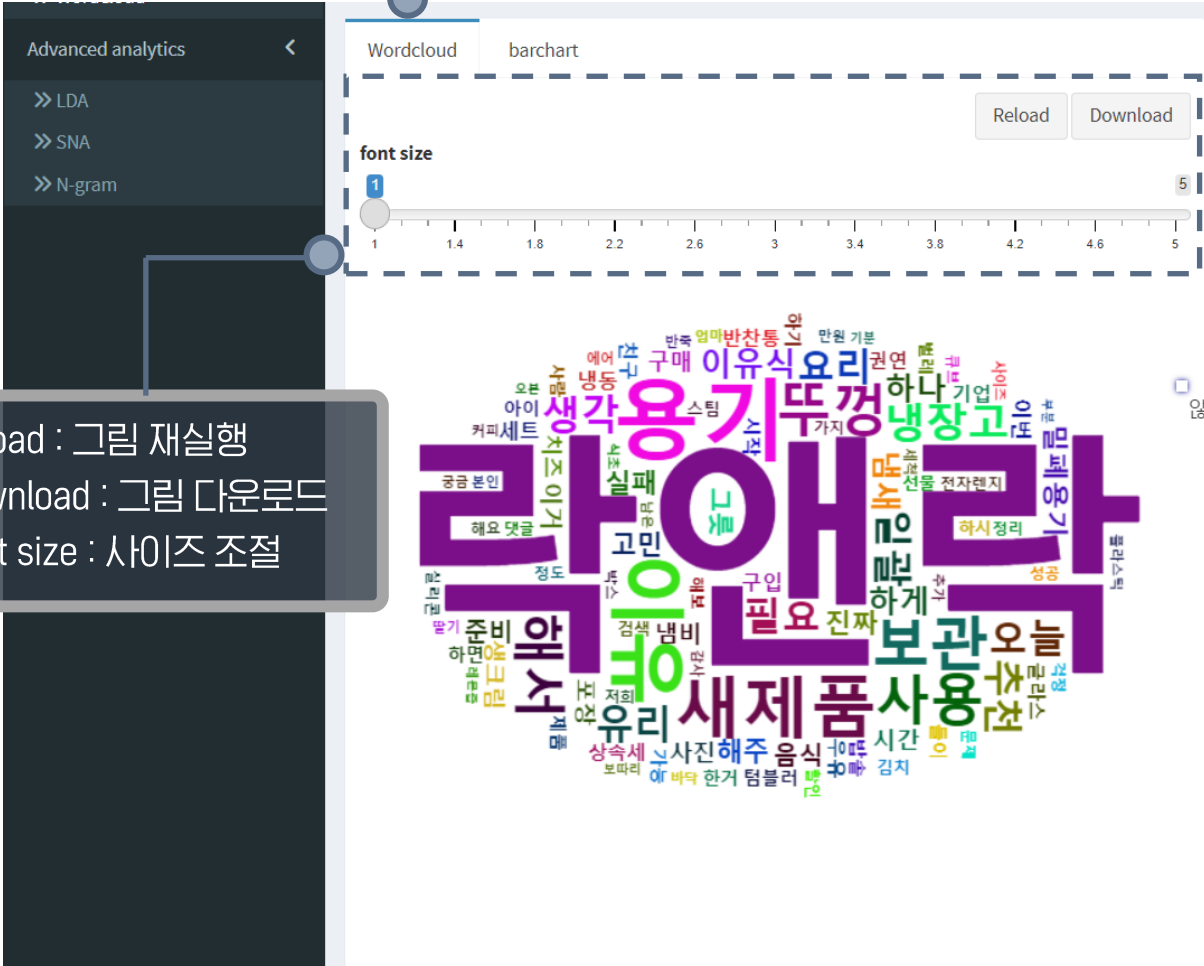
Download Next

Basic analytics >> Wordcloud

빈도 상위 100개의 단어에 대한
 - Wordcloud : 워드클라우드 결과
 - barchart : 바차트 결과

Datatable
 빈도 상위 100개의 단어에 대한 데이터 테이블
 * Remove 버튼을 누르면 단어가 제거되며,
 이후 심화분석 과정에도 동일하게 제거됩니다.

Reload : 그림 재실행
 Download : 그림 다운로드
 Font size : 사이즈 조절



Datatable

No.	Word	Freq	Remove
1	락앤락	353	Remove
2	용기	124	Remove
3	이유	108	Remove
4	새제품	95	Remove
5	보관	82	Remove
6	뚜경	79	Remove
7	사용	77	Remove
8	해서	75	Remove
9	냉장고	59	Remove
10	생각	55	Remove

Showing 1 to 10 of 100 entries

Previous 1 2 3 4 5 ... 10 Next

Basic analytics >> Wordcloud

Datoz

- Setting <
- >> File upload
- >> Data processing
- Basic analytics <
- >> Time-series charts
- >> **Wordcloud**
- Advanced analytics <
- >> LDA
- >> SNA
- >> N-gram

Word Count

Count

 Select


Download
Next

Wordcloud barchart

Reload Download

font size

1 5



Datatable

Next를 누르면,
Advanced analytics (심화분석)의
LDA 목차로 넘어갑니다.

No.	Word	Freq	Remove
1	이유	353	Remove
2	용기	124	Remove
3	이유	108	Remove
4	새제품	95	Remove
5	보관	82	Remove
6	뚜껑	79	Remove
7	사용	77	Remove

Advanced analytics >> LDA(tab: Evaluation indicator)

해당 Select 버튼을 누르면,
LDA 탭으로 자동 이동하며, 버튼을 누른 이후에는
자유롭게 탭을 이동하여 관련 결과를 확인할 수 있습니다.
* 해당 버튼을 클릭하여 LDA, SNA, N-gram 분석이
진행되므로, Select 버튼을 누르지 않고 다른 심화분석 확인은 불가능합니다.

Datoz

- Setting <
- >> File upload
- >> Data processing
- Basic analytics <
- >> Time-series charts
- >> Wordcloud
- Advanced analytics <
- >> LDA
- >> SNA
- >> N-gram

LDA(Latent Dirichlet Allocation)

추출한 문서에 담긴 단어들의 주제(토픽)를 추출하는 기법.
단순 주제만 분류해주는 것이 아니라 주제에 포함되는 키워드들을 보여 주기 때문에 그 키워드들로 해당 주제를 해석하고 정의할 수 있음.

Evaluation indicator LDA Detail

Number of Topics

 Select

[토픽 선정지표]

위의 패널에 있는 Arun2010과 CaoJuan2009는 작을수록
아래의 Griffiths2004와 Deveaud2014는 클수록 최적의 값으로 해석
(단, 네 값의 측도가 모두 최적이 되는 상황이 없으므로
해석의 용이성을 감안하여 토픽의 개수를 선정)

그래프를 확인하고,
선정지표에 따라 최적의 토픽 개수를
선정합니다.

Advanced analytics >> LDA(tab: LDA)



Number of terms to display
토픽별 포함되는 단어개수 조정

Number of terms to display: 20 30 40

Selected Topic: 0 Previous Topic Next Topic Clear Topic

Slide to adjust relevance metric:⁽²⁾ $\lambda = 1$ 0.0 0.2 0.4 0.6 0.8 1.0

Intertopic Distance Map (via multidimensional scaling)

Top-30 Most Salient Terms⁽¹⁾

Term	Count
락앤락	550
응기	150
유리	100
이유	100
해서	100
오늘	100
생각	100
사진	100
고민	100
보관	100
시각	100

HTML형태로 구현된 LDA 분석결과 화면
각 토픽별로 마우스를 올려놓으면
결과를 확인 할 수 있습니다.

Advanced analytics >> LDA(tab: Detail)

토픽별 단어 리스트 차트와 테이블
 토픽에 대한 영향이 높은 단어순으로 내림차순 나열
 전체 토픽 또는 각 토픽별 단어리스트 차트를
 확인 할 수 있습니다.

Datoz

- Setting <
- >> File upload
- >> Data processing
- Basic analytics <
- >> Time-series charts
- >> Wordcloud
- Advanced analytics <
- >> **LDA**
- >> SNA
- >> N-gram

LDA(Latent Dirichlet Allocation)

추출한 문서에 담긴 단어들의 주제(토픽)를 추출하는 기법.
 단순 주제만 분류해주는 것이 아니라 주제에 포함되는 키워드들을 보여주기 때문에 그 키워드들로 해당 주제를

Evaluation indicator
LDA
Detail

전체

전체

Topic.1

Topic.2

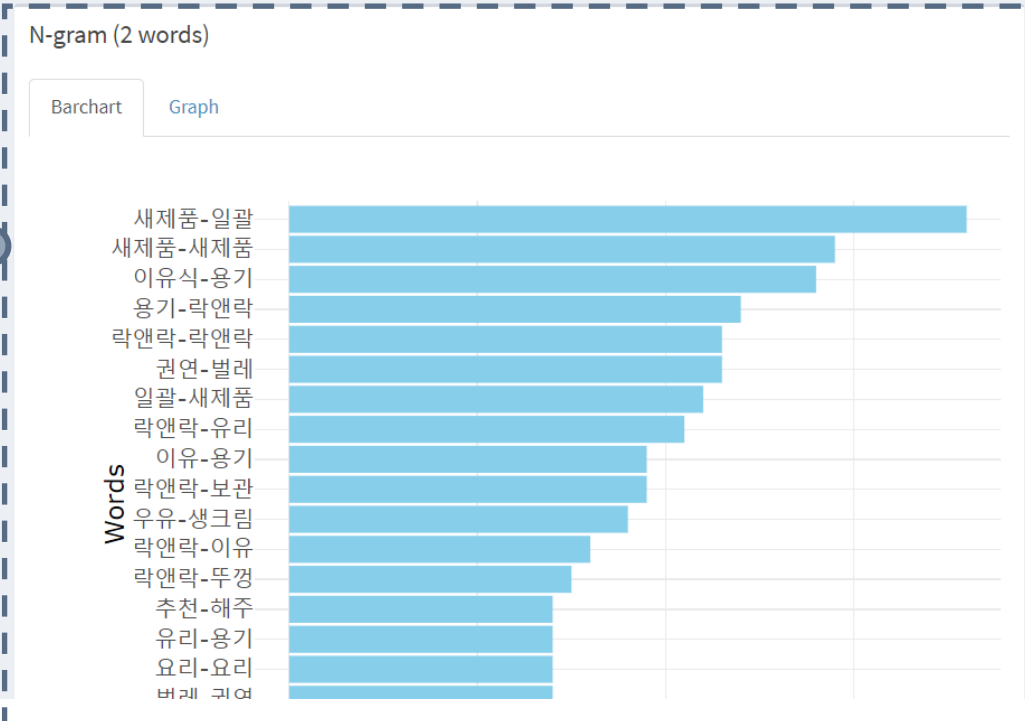
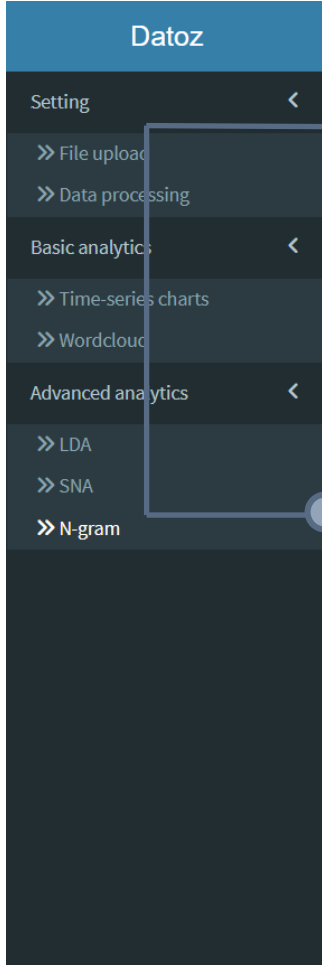
Topic.3

Topic.1	Topic.2	Topic.3
이유	락앤락	해서
보관	용기	유리
생각	고민	오늘
냉장고	사용	뚜껑
락앤락	구매	사진
필요	시간	하계
하나	한거	하면
그릇	구입	해보
추천	가능	검색

Advanced analytics >> N-gram(tab: Barchart)

N-gram (2 words)
 2개의 연속적인 단어의 나열을 하나의 묶음으로 간주하여
 이를 Barchart와 Graph로 시각화
통계학 기반의 언어 모델 중 하나로 문장 내 앞서 등장한 단어를 기반으로 이어서 등장할 적절한 단어를 예측하는 분석.
 n개의 연속적인 단어의 나열을 하나의 묶음(-token)으로 간주.
 * 아래의 N-gram (3 words)는 3개의 연속적인 단어의 나열을 묶음으로 간주

Datatable
 연속적인 단어의 나열을 하나의 묶음으로 간주하여
 Words로 표현하고
 각각의 Words에 대한 빈도순으로 나열한 데이터테이블



Rank	words	frequency
1	새제품-일괄	36
2	새제품-새제품	29
3	이유식-용기	28
4	용기-락앤락	24
5	권연-벌레	23
6	락앤락-락앤락	23
7	일괄-새제품	22
8	락앤락-유리	21
9	락앤락-보관	19
10	이유-용기	19

Advanced analytics >> N-gram(tab: Graph)

Frequency filter

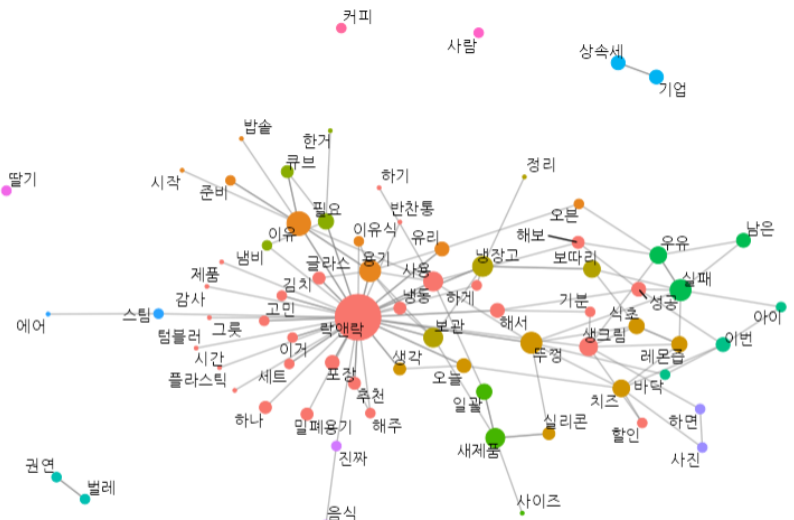
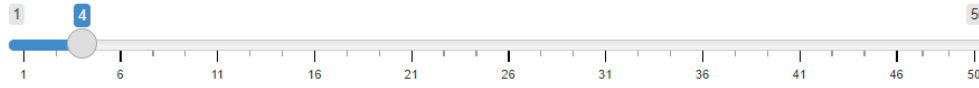
frequency의 최소값을 지정

ex) frequency filter 값이 10일 경우,
N-gram frequency 값이 10 이상인 단어 나열 묶음이 그래프로 그려집니다.

Barchart Graph

동시 출현 빈도가 높은 단어들 중심이 되어 네트워크 형태로 표현,
단어들 간의 연관(동시 출현 빈도)이 높을 경우 같은 색상으로 표현되며 개별 단어의 빈도가 높을수록 노드의 크기가 커짐

frequency filter



Datatable

Download

Rank	words	frequency
1	새제품-일괄	36
2	새제품-새제품	29
3	이유식-용기	28
4	용기-락앤락	24
5	권연-벌레	23
6	락앤락-락앤락	23
7	일괄-새제품	22
8	락앤락-유리	21
9	락앤락-보관	19
10	이유-용기	19

Showing 1 to 10 of 153 entries

Next

Graph

동시 출현 단어 빈도 기준으로
네트워크 형태로 도식화하여 구현

E n d O f D o c u m e n t

Contact Us

Website URL <http://bigdata.emforce.co.kr>

T 02. 6177. 1871 eMAIL bigdata@emforce.co.kr