**커피 데이터 분석 보고서**

**서론**

최근 들어 커피에 대한 대중의 관심이 크게 증가했지만, 어떤 커피를 선택해야 할지 몰라 많은 소비자들이 고민하고 있습니다. 이에 따라, 로스팅 수준과 산지에 따른 커피의 특성을 데이터 분석과 시각화를 통해 탐구하고, 좋은 커피를 선택할 수 있는 기준을 제공하고자 합니다. 또한 로스팅 수준별로 가성비 좋은 커피를 추천하여 실질적인 구매 도움을 제공하고자 합니다.

**본론**

1. **로스팅 수준별 기초 통계량 분석**

도표, 직사각형, 평면도, 텍스트이(가) 표시된 사진

자동 생성된 설명

로스팅 수준에 따른 평점의 기초 통계량은 다음과 같습니다:



* **평균 평점**은 Light 로스팅(93.52)이 가장 높았고, Dark 로스팅(88.20)이 가장 낮았습니다.
* **분산**은 Dark와 Medium-Dark 로스팅이 상대적으로 높아 평점 차이가 클 가능성을 보였습니다.

**2. ANOVA 테스트**

일원분산분석(One-Way ANOVA)

정의: ANOVA는 세 개 이상의 그룹 간 평균 차이가 통계적으로 유의미한지 확인하는 데 사용됩니다. 여기서는 로스팅 수준(Light, Medium, Dark)을 그룹으로 설정하여 평점 평균 차이를 검증했습니다.

* 가정
* 정규성 가정:

각 그룹의 데이터가 정규분포를 따른다고 가정합니다.

* 분산 동일성 가정:

모든 그룹의 분산이 동일하다고 가정합니다.

* 수행 절차
* 집단별 데이터 분리
* 로스팅 수준별로 평점을 그룹화합니다.
* F-Statistic 계산:

집단 간 변동(Between-Group Variance)과 집단 내 변동(Within-Group Variance)을 비교하여 F-Statistic을 계산합니다.

* P-Value 계산

귀무가설(모든 그룹의 평균이 동일하다)을 검증하기 위해 P-Value를 계산합니다.

로스팅 수준 간의 평점 차이를 확인하기 위해 ANOVA 테스트를 수행한 결과는 다음과 같습니다.

* **F-statistic**: 47.91
* **p-value**: 0.0000

p-value가 0.05보다 작으므로, 로스팅 수준에 따라 평점의 유의미한 차이가 있다고 결론지을 수 있습니다.

**3. 로스팅 수준별 가성비 좋은 커피 추천**

로스팅 수준별로 가성비(평점/가격)가 가장 높은 커피를 다음과 같이 추천합니다



**결론**

이번 분석을 통해 다음과 같은 인사이트를 얻을 수 있었습니다:

1. 로스팅 수준에 따라 평점의 차이가 유의미하게 존재합니다. 특히 Light 로스팅은 평균 평점이 가장 높았으며, Dark 로스팅은 상대적으로 낮았습니다.
2. 가성비 분석 결과, **Medium-Light 로스팅의 Karen J Kona Red Bourbon** 커피가 가장 높은 가성비를 보여주었습니다.