**Growth Hackers 8기 지원서**

안녕하세요, Growth Hackers 8기에 관심을 갖고 지원해 주셔서 대단히 감사합니다.

**지원 방식은 다음과 같습니다. 1) 지원서(Word), 2) Quest 풀이 설명 파일(Word/PPT/한글), Quest 해결을 위해 사용한 3) 소스 코드 파일(R/ipynb/py)**을 함께 압축하여 **‘recruit@ghsnu.com’**으로 보내주시기 바랍니다. (지원서, 압축 파일 및 이메일 제목: **‘(지원자 이름)\_GH 8기 지원서’**)

지원서 관련 문의사항은 동일한 메일 혹은 카카오 플러스 친구 @ghsnu8th으로 문의해 주시면 감사하겠습니다.

리크루팅 관련 주요 일정은 다음과 같습니다.

|  |  |  |  |
| --- | --- | --- | --- |
| **설명회** | **서류 지원 마감** | **면접** | **OT** |
| 9월 1일(화)  오후 7시  비대면 | 9월 3일(목)  오후 6:00 | 9월 5일(토)  ~ 9월 8일(화) | 9월 12일(토)~9월 13일(일)  \*코로나 확산 상황에 따라 취소될 수 있음. |

**지원 시 유의사항 (지원 전 꼭 읽어 주세요!!)**

1. 제출 기한은 9월 3일 목요일 **‘오후 6시’** 입니다.
2. Growth Hackers는 학회 특성상 기초 프로그래밍 및 데이터 분석 소양을 확인하기 위한 Quest가 포함되어 있습니다. 지원서 작성에 넉넉한 시간을 확보하시어 제출 기한을 놓치지 않으시길 바랍니다.
3. Growth Hackers는 매주 화(오후 7:30), 목(오후 7:30), 토(오후 1:30) 주 3회 정기 Session을 진행하고 있습니다. 그 외에도 팀 별 과제, Quest, 등 추가적인 시간을 할애해야 하는 점 숙지하시기 바랍니다. 추가로 에듀 세션이 진행되는 9 월 중순~ 10 월 초에는 월(오후 7:30), 금(오후 7:30)에도 세션이 예정되어 있습니다. 정확한 일자는 ghsnu.com 의 공지사항을 확인해주시면 감사하겠습니다.
4. Growth Hackers는 다른 학회원들에게 피해가 가지 않도록 하기 위해 결석과 지각에 대한 벌금을 부여하고 있습니다. 또한, 누적 결석 횟수가 일정 횟수를 초과할 경우 Growth Hackers 활동을 함께하기 어렵습니다.
5. 본인이 배우고 싶은 언어나 프로그램을 공부하지 못할 수도 있습니다. 소모임을 구성하여 다른 언어와 프로그램도 다룰 수 있지만, 주 커리큘럼에서는 Python을 공부하게 됩니다.

**1. 기본 인적 사항 (아래 링크에서 면접 가능 시간대를 작성해주시기 바랍니다.)**

|  |  |  |  |
| --- | --- | --- | --- |
| **성명** | 황창현 | **생년월일** | 981203 |
| **소속 대학교** | 세종대학교 | **입학연도 /**  **수료 학기 수/**  **남은 학기 수** | 2017/  3/  5 |
| **주전공** | 전자정보통신공학과 | **복/부/연합 전공** | 비즈니스애널리틱스(연계) |
| **연락처** | 01077023047 | **E-mail** | ckdckd01@gmail.com |

**\* 링크 :** [**https://forms.gle/JpeFyLQBV22yRd8R8**](https://forms.gle/JpeFyLQBV22yRd8R8)

**2. 추가 사항 (해당 사항이 없을 시 기입하지 않으셔도 됩니다.)**

|  |
| --- |
| **학교 혹은 개인적으로 Data Science / Business 관련하여 들은 과목이나 강좌 (과목명, 수강 일시)** |
| 그로스해킹(데이터와 실험을 통해 성장하는 서비스를 만드는 방법-인프런)20200427 시작.20200709수료.  구글 애널리틱스(구글 애널리틱스 아카데미) gaiq 자격증 20210526 만료.  빅데이터 분석실무 2급 (한국정보인재개발원-세종대학교) 2020723~20200724 끝. |
| **다룰 수 있는 프로그래밍 언어의 종류, 파이썬 라이브러리 및 그 수준에 대해 간단하게 작성해주세요**  **(평가 대상이 되지 않는 문항으로써 작성하지 않으셔도 어떠한 패널티도 주어지지 않습니다).** |
| python(numpy - 소, pandas - 중 , scikit-learn - 소, lightgbm - 중, bayes\_opt - 소)  R(소) |
| **예정된 학회 기간 중 참여하지 못하는 날이 있다면 미리 적어주시기 바랍니다.**  **미리 알려주신 이외의 날짜에 중대한 사유로 불참하는 경우가 아니면, 벌금 및 벌점이 부과될 수 있습니다.** |
| 토요일 (공모전-빅콘테스트 팀원들과의 미팅시간이 주로 토요일 오전에 잡혀서 우선 9월 28일 까지 참여가 어려울 수 있습니다. 그 외 토요일은 아주 가끔 할머니네 가족들이랑 같이 가고 해서 못갈 수도 있습니다.) 목요일 수업이 6시에 끝납니다. 화요일 수업이 7시 30분에 끝납니다. 지각 할 것 같습니다. |
| **학회에 들어오게 된다면 함께 프로젝트를 하고 싶은 기업, 혹은 프로젝트 주제가 있다면 적어주시기 바랍니다. (본 문항은 프로젝트를 함께 할 기업 소싱 참고용으로만 사용되며, 평가 대상이 되지 않는 문항으로 작성하지 않으셔도 어떠한 패널티도 주어지지 않습니다.)** |
| 나동빈 블로그[<https://ndb796.tistory.com/>] 중 선형회귀 프로젝트를 보면 배추값예측하는 사이트 만드는게 있는데, 금융데이터에 관심이 많습니다. 플라스크를 통해 웹서비스를 만들고 배포는 어떻게 하는지 아직 공부못했지만, 내용은 매출 관련 금융데이터를 받아 매출예측을 하여 사이트로 보여주는 사이트를 만들고 싶습니다. |

**3. 자기소개서**

|  |
| --- |
| **1. 수업이나 대외활동, 인턴 등을 통해 데이터 분석 / Business 관련하여 진행하였던 프로젝트 중 가장 인상 깊었던 경험에 대해 적어주세요. 자신이 수행했던 역할은 무엇이었는지, 역경이 있었다면 어떻게 극복하였는지 상세히 적어주세요. (해당 주제 관련 프로젝트 경험이 없다면 다른 주제의 프로젝트나 팀플 등에 대해 적어주시면 됩니다.)** |
| 최근 데이콘 서포터즈 2기를 마쳤습니다. 이번 활동 내용은 이미 마감된 경진대회 하나의 우승자코드를 필사하는 활동이고, 저는 팀장으로 활동했습니다. 대회는 **상점매출예측** 대회였고, 전체적인 진행 및 커리큘럼 개요는 데이콘 측에서 제공해줬기 때문에 큰 어려운 부분없이 진행했습니다. 역경이 있었다면, 그 당시 저는 파이썬에 대한 이해도가 높지 않았고, kaggle 필사도 어려워했습니다. 그래서 같이 서포터즈하는 다른 분들께도 물어보고 진행하는 매니저님한테도 물어보며 해결해갔습니다. 어떤 질문을 받을 지 모르니 항상 ppt로 커리큘럼의 모든 부분을 1주,2주전에 공부하여 저장해놓았습니다. 막히는 일 없이 잘 풀어갈 수 있었습니다. 또, 팀원들과의 커뮤니케이션과 협업에 문제가 있었습니다. 왜냐하면 모든 스터디는 온라인으로 진행했기에 직접적인 케어가 부족했습니다. 조금이라도 스터디원의 질문과 공부를 위해 slack커뮤니케이션을 통해 채널을 관리하여 스터디내용을 공유했고, google meeting을 통해 일주일에 한번 꼭 진행상황 과 발표를 진행하여 이해여부를 확인했습니다. 마지막에 제출까지는 같이 못했지만, 데이콘 코드공유에 공유도 하고 점수도 성공적으로 끝마쳤습니다. |
| **2. 이 외에도 지금까지 가장 열정을 가지고 몰두했던 일이 있었다면 무엇인가요? 그로 인해 무엇을 성취할 수 있었나요? (1번 답변과 중복되지 않는 경험에 대해 서술해주세요.)** |
| 또 다른 데이콘 대회 중 **공공데이터를 활용 온도추정 ai 경진대회**가 있는데, 마감 된 대회지만, 다시 프로젝트 진행하여 public score 4.1(상위 약8%) 까지 했습니다. 맨 처음 deep learning lstm 기법으로 무작정 돌려봤던 대회 입니다. 이번엔 lstm이 아닌lightgbm 기법으로 진행했고, bayes\_opt를 통한 하이퍼파라미터 튜닝까지 진행했습니다. 이번 프로젝트를 통해 lightgbm 파라미터에 대해서 이해했습니다.  최근, 홈페이지를 만들었습니다. <https://handmades.netlify.app/>  원래 블로그를 운영했다가 컴퓨터도 복구되고 유지하기 어려워질것같아 파이썬으로 정적홈페이지를 만들었습니다. Flask 에 대해서 이해했다는 생각을 했고, 배포같이 막히는 부분을 어떻게 해결해야할지 알았습니다. |
| **3. 학회에서 본인이 기여할 수 있는 부분이 무엇이라고 생각하십니까? (개인의 경험 및 사례를 들어서 서술해주세요.)** |
| 저를 포함해 많은 분들은 눈에 보이는 결과물을 원할 것입니다. 하지만 어떻게 해야할 지 막막해 하시는 분들이 많습니다. 인터넷강의를 봐도 따라만 하는것이고 결국 자신만의 것으로 만들지 못합니다. 저는 제가 공부한 방법을 얘기하려고 합니다. lightgbm 몰랐습니다. kaggle titanic 부터 house price project, predict future sales 까지 좋은 부분이 있다면 가져와서 몰라도 계속 썼습니다. def lgb\_train(params.....): .... 길다고 생각한 코드는 10번 반복하면 50줄 코드를 순식간에 쓸 것이고, 이제 쓰였던 코드들의 패턴, 문법, 이유 등이 눈에 들어옵니다. 그때 구글에 검색을 합니다. lightgbm lambda\_l1 parameter 단어 위주로 검색해도 좋고 문장으로 검색해도 좋습니다. 하나씩 자기것으로 만드는것입니다. 저는 검색하는 방법, 찾는 방법, 무작정 써보는 공부방법을 학회에 기여할 수 있다고 생각합니다. |
| **4. 데이터를 활용해서 해결해보고 싶은 문제 혹은 프로젝트가 있다면 무엇인가요?** |
| 1. 나동빈 블로그[<https://ndb796.tistory.com/>] 중 선형회귀 프로젝트를 보면 배추값예측하는 사이트 만드는게 있는데, 금융데이터에 관심이 많습니다. 플라스크를 통해 웹서비스를 만들고 배포는 어떻게 하는지 아직 공부못했지만, 내용은 매출 관련 금융데이터를 받아 매출예측을 하여 사이트로 보여주는 사이트를 만들고 싶습니다.  2. 데이터 베이스를 공부하여 큰 데이터를 효과적으로 다루는 프로젝트를 하고 싶습니다. |

**Growth Hackers 8기 Quest**

**※ 퀘스트에는 [알고리즘]과 [데이터분석] 영역이 있으며 각 영역은 2문항으로 구성되어 있습니다. 정답 여부보다는 풀이 과정을 보기 위한 문제이므로 부담 갖지 않으셔도 됩니다.**

**※ 퀘스트 풀이 설명 파일은 Word, 한글, PPT 등을 통해 자유롭게 작성하여 제출해주시면 됩니다. 다만, 문제에 대한 코드와 답안 및 실행결과 창의 스크린샷을 포함하여 작성해주시기 바랍니다.**

**※ 코딩이 처음이신 경우, Codecademy의 기초 Python 강좌 혹은 Jump to Python(박응용 著)을 통해 Python 학습하신 후 문제를 해결하실 수 있습니다. (Jump to Python 링크:**[**https://wikidocs.net/book/1**](https://wikidocs.net/book/1)[**)**](https://wikidocs.net/book/1))

**※ 퀘스트와 관련된 문의 사항은** [**recruit@ghsnu.com**](mailto:recruit@ghsnu.com) **또는 카카오 플러스 친구 @ghsnu8th 친구 추가 후 1:1 문의로 보내주시면 됩니다.**

**[알고리즘] 영역**

※ **이 영역은 Python 기초 소양을 측정하기 위한 것이므로 Python을 활용하여 해결해주세요.**

1. ‘체인 소수’는 한 숫자를 연속적으로 잘라 얻을 수 있는 값들과 자기 자신이 모두 소수인 숫자를 일컫는다. 예를 들어 ‘23’에서 얻을 수 있는 값은 ‘2’, ‘3’, ‘23’으로 ‘23’은 체인 소수로 볼 수 있다. 이에 반해, ‘127’로 만들 수 있는 값은 ‘1’, ‘2’, ‘7’, ‘12’, ‘27’, ‘127’로 ‘127’은 체인 소수가 아니다. **입력 받은 숫자가 체인 소수인지 판별하는 함수를 작성하라.**

(e.g.) input=23 → output=True / input=29 → output=False

1. 신입생이 학교에 입학할 때 부여되는 ‘학번’은 일정한 규칙을 가지고 만들어진다. 학번은 가령 ‘2020-12345’의 형태와 같이 총 9자리의 숫자로 구성되는데, 처음 네 자리는 입학년도로 설정된다. 그 다음 숫자는 학부/석사/박사의 경우 각각 1/2/3이 부여되며, 마지막 네 자리는 랜덤한 값으로 정해진다. 이 때, 새로운 학생이 입학할 때마다 딕셔너리에 학번을 키(key), 학생의 이름을 밸류(value)로 저장하는 함수를 만들고자 한다. 위 규칙에 따라 **‘입학년도’, ‘학부/석사/박사 여부’, ‘이름’을 입력받고 학번을 생성하여 이를 딕셔너리에 학번을 키(key), 이름을 밸류(value)로 저장하는 함수를 작성하라.** (단, 중복되는 학번이 만들어져서는 안 된다.)

(e.g.) input='2019'.'학부','김그로스' → output={'2019-19414' : '김그로스'}

/ input='2014','박사','최해커스' → output={'2019-19414' : '김그로스', '2014-32768' : '최해커스'}

**[데이터 분석] 영역**

※ **이 영역은 기초 통계 이해를 측정하고 논리적 분석과정을 파악하기 위한 것이므로 자유로운 툴과 방식으로 해결해주세요. 툴의 종류는 평가에 영향을 주지 않습니다.**

※ **제공된 데이터셋인 1. 앱스토어 데이터 ‘app\_store.csv’ 와 2. 음원 정보 데이터 ‘music\_db.csv’ 중 본인이 더 자신이 있는 유형의 데이터셋과 문제를 반드시 하나만을 선택하여 풀어주세요. 두 유형을 모두 풀이에 적어 제출하실 경우 오히려 채점에 불이익이 있을 수 있습니다.**

※ **본 문제에서 가정한 상황은 모두 허구이며, 실제로 기업, 단체와는 아무런 상관이 없습니다.**

1. 앱스토어 데이터

다음의 app\_store.csv 데이터는 신 사업으로 어플 추천 시스템을 개발하는 스타트업인 G사가 다양한 어플에 대한 특성 정보와 사용자 평가 정보를 확보하여 정리한 데이터이다.

※ Rating은 평점, Reviews는 리뷰의 수, Installs는 설치된 수, Size는 앱의 용량, 그리고 Price는 앱의 가격이다.

1 - 1 ) Rating 변수와 상관관계를 분석할 수 있는 변수를 찾고, 그들이 어떠한 관계를 갖는지 이유와 함께 설명해주세요. (**시각 자료 및 코드를 제외한 답안의 길이는 한 페이지 이내로 제한해주세요.)**

1 - 2 ) 250번 (Five Nights at Freddy's 4), 532번 (PAW Patrol Pups to the Rescue), 1664번 (My Town : Farm) 어플을 사용하는 고객이 만족할 만한 어플을 추천해주는 전략을 구체적인 추천 목록과 함께 제시해주세요. 단, 이 고객의 잔여 용량은 500M(500,000k), 예산 범위는 20달러로 추천하는 어플들의 용량과 가격 총합이 제시된 제한 범위를 넘어서는 안 됩니다. / 고객이 만족할 만한 어플에 대한 기준은 자유롭게 설정하시고 그 근거를 설명해주시면 됩니다. (**시각 자료 및 코드를 제외한 답안의 길이는 두 페이지 이내로 제한해주세요.)**

2. 음원 정보 데이터

다음의 music\_db.csv 데이터는 음원 스트리밍 기업인 H사가 보유하고 있는 음원의 여러 특성 정보들을 정리한 통합 데이터이다.

※ music\_num은 변수가 아닌 곡의 번호, popularity는 곡의 인기 정도, duration\_ms는 곡의 길이, key는 곡의 음계, major는 곡의 장단조 여부, time\_signature은 곡의 박자 그리고 valence는 곡의 분위기를 수치화한 변수이며 기타 변수들 역시 곡들의 여러 특성을 수치화한 변수이다.

2 - 1 ) Loudness 변수가 다른 모든 변수들과 독립적이라고 말할 수 있는지에 대한 본인의 생각을 그 이유와 함께 구체적으로 설명해주세요. (**시각 자료 및 코드를 제외한 답안의 길이는 한 페이지 이내로 제한해주세요.)**

2 - 2 ) H사는 고객들의 취향에 맞는 음악 추천을 위해 주어진 데이터의 음악들을 그룹지어 나누고자 합니다. 주어진 음악들의 다양한 특성 데이터들을 활용하여 어떤 기준으로 몇 개의 그룹으로 나누는 것이 가장 좋을지 구체적인 그룹화 결과와 함께 제시해주세요. / 그룹화에 대한 기준은 자유롭게 설정하시고 그 근거를 설명해주시면 됩니다. (**시각 자료 및 코드를 제외한 답안의 길이는 두 페이지 이내로 제한해주세요.)**