

# Mental Health In Company

라이언조

경영정보학과 이건우

산업경영공학과 이상인

컴퓨터 공학과 김하은

# 목차

---

01

목표 및 데이터 설명

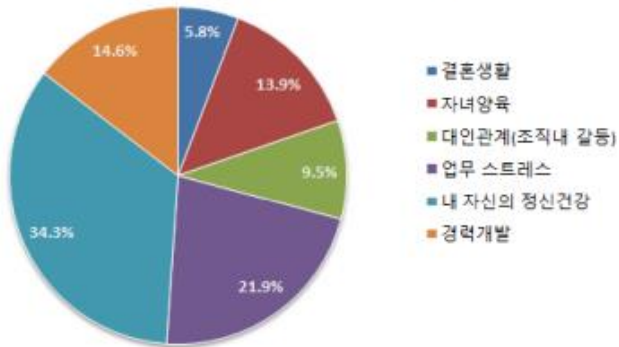
02

데이터 분석 과정  
및 결과

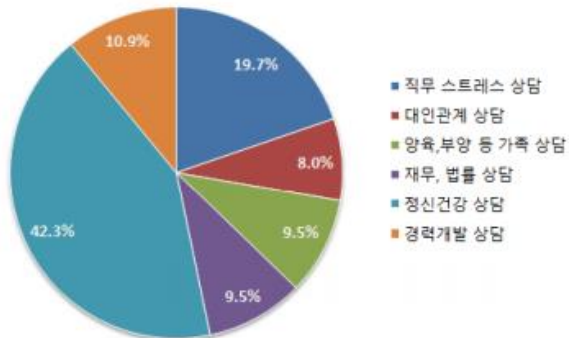
03

결론 종합

## 직장인에게 가장 고민되는 문제

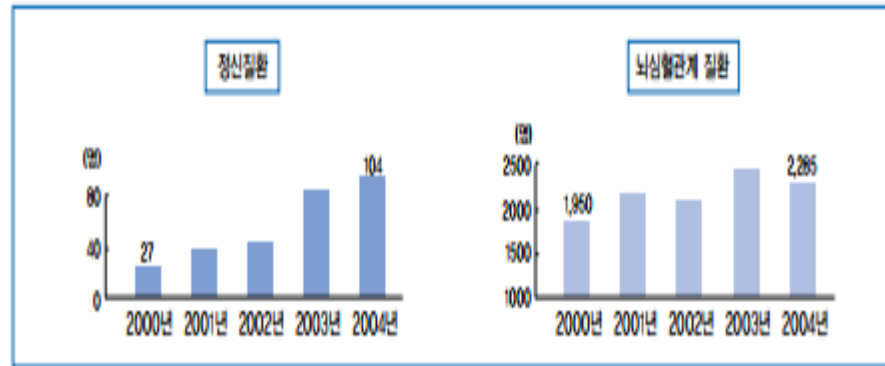


## 직장인에게 가장 필요한 상담 분야



출처 : 복지기업 이지엘페어 설문조사

(그림 1) 스트레스 관련 산업재해 증가(산재승인 건수)



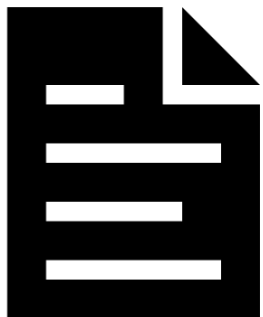
자료: 근로복지공단



1. 고용주와 고용인의 관계 및 회사의 환경이 직원의 정신 건강에 어떠한 결과를 주고, 직원들의 대처 방안에 영향을 주는지 확인
2. 고용주의 정신 건강과 관련된 복지제도가 고용인에게 어떠한 영향을 주는지 그 영향과 중요도를 확인



1. 고용인의 정신 건강을 위한 고용주와 고용인의 관계 형태 제시
2. 고용주가 제공하는 정신 건강 복지제도의 형태 제시



Mental Health in Tech Survey  
(2014년 미국 IT회사들을 대상으로 한 설문조사)

Random Forest

Association  
Analysis

Logistic  
Regression

- state 변수를 제외한 모든 변수로 RF를 진행

```
##          OOB estimate of  error rate: 52%
```

```
## Confusion matrix:
```

```
##      No Yes class.error
```

```
## No   185  10  0.05128205
```

```
## Yes  289  91  0.76052632
```

No의 잘못 예측 비율 : 5%  
Yes의 잘못 예측 비율 : 76%

- TuneRF 함수를 이용하여 오류 확률이 최소가 되는 최적의 parameter 값 찾기

```
## mtry = 4  OOB error = 46.96%
## Searching left ...
## mtry = 8   OOB error = 47.65%
## -0.01481481 0.05
## Searching right ...
## mtry = 2   OOB error = 45.04%
## 0.04074074 0.05
```



**mtry = 2, ntreeTry = 50 일 때 오류 확률 45.04%로 최저**

```
##          OOB estimate of  error rate: 35.3%
```

```
## Confusion matrix:
```

```
##      No Yes class.error
```

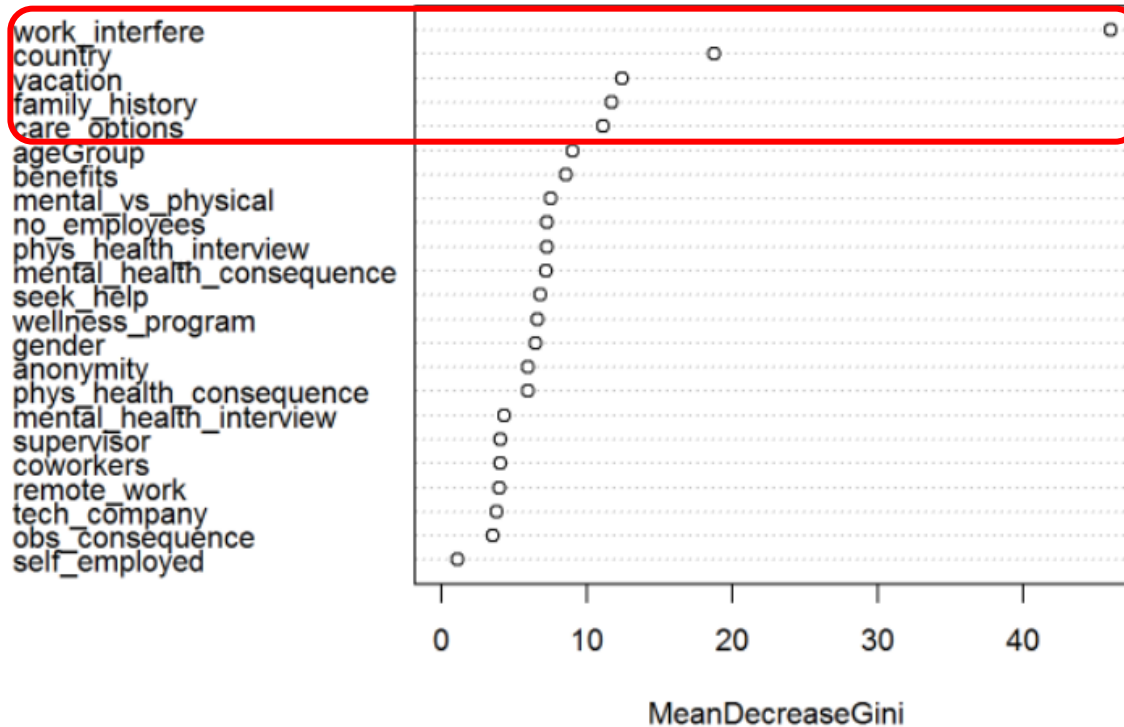
```
## No   153  42  0.2153846
```

```
## Yes  161 219  0.4236842
```



**mtry = 2, ntreeTry = 50로 다시 RF 한 결과,  
오류 확률이 35.3%로 전보다 많이 낮아짐**

Variable Importance



- work\_interfere가 가장 상위에 있음
- 그 다음으로 country, vacation, family\_history, care\_option 순으로 중요도가 나열됨

- RHS= "treatment=No"로 고정하였을 때,

```
asso0 <- apriori(sur, parameter = list(minlen = 3, support = 0.1, confidence = 0.1),
  appearance = list(rhs = "treatment=No", default = "lhs"),
  control = list(verbose = F))
```

```
> inspect(asso0[1:5])
```

lhs	rhs	support	confidence	lift
[1] {family_history=No,work_interfere=Never}	=> {treatment=No}	0.1529284	0.9038462	2.495048
[2] {country=United States,work_interfere=Never}	=> {treatment=No}	0.1030369	0.8715596	2.405922
[3] {work_interfere=Never,anonymity=Don't know}	=> {treatment=No}	0.1301518	0.8695652	2.400417
[4] {work_interfere=Never,wellness_program=No}	=> {treatment=No}	0.1149675	0.8760331	2.418271
[5] {work_interfere=Never,supervisor=Yes}	=> {treatment=No}	0.1344902	0.8794326	2.427655

- ▶ 전반적으로 작업환경에 있어 **정신 건강 상태가 방해가 된다고 느끼지 않을 때**(work\_interfere=Never), 정신 건강에 대한 치료법을 찾지 않음
- ▶ 특히 **가족 병력이 없을 때** 그 경향이 더 강하게 나타남

- RHS= "treatment=Yes"로 고정하였을 때,

```
> inspect(asso1[1:5])
```

lhs	rhs	support	confidence	lift
[1] {work_interfere=Often,coworkers=Yes}	=> {treatment=Yes}	0.1008677	0.8691589	1.362865
[2] {work_interfere=Often,mental_health_interview=No}	=> {treatment=Yes}	0.1019523	0.8623853	1.352244
[3] {coworkers=Yes,obs_consequence=Yes}	=> {treatment=Yes}	0.1073753	0.8048780	1.262071
[4] {mental_health_interview=No,obs_consequence=Yes}	=> {treatment=Yes}	0.1138829	0.7608696	1.193064
[5] {self_employed=No,obs_consequence=Yes}	=> {treatment=Yes}	0.1008677	0.7440000	1.166612

- ▶ 작업 환경에 있어 **정신 건강 상태가 꽤 방해된다고 생각될 때**(work\_interfere=Often), 정신 건강에 대한 치료법을 찾으려 함
- ▶ 특히 정신 건강에 대한 문제를 **주변 동료들과 논의할 수록** 그 경향이 강하게 나타났고, 주변에서 **부정적인 결과에 대해 들을 때**(obs\_consequence=Yes) 치료법을 찾으려 함



country = "United States"이고

- RHS= "treatment=No"로 고정하였을 때,

```
> inspect(asso_us_0_2[1:5])
```

lhs	rhs	support	confidence	lift
[1] {family_history=No,mental_health_interview=Maybe}	=> {treatment=No}	0.05026930	0.6363636	1.947552
[2] {supervisor=yes,mental_health_interview=Maybe}	=> {treatment=No}	0.05385996	0.4411763	1.330194
[3] {phys_health_consequence=No,mental_health_interview=Maybe}	=> {treatment=No}	0.05026930	0.4516129	1.382134
[4] {gender=Male,mental_health_interview=Maybe}	=> {treatment=No}	0.05385996	0.4838710	1.480858
[5] {coworkers=yes,mental_health_interview=Maybe}	=> {treatment=No}	0.05565530	0.4428571	1.355338

- ▶ 가족 병력이 없고(family\_history=No), 잠재적 고용주와 면접에서 정신건강에 대해 논의하는데 확실하지 않은(mental\_health\_interview=Maybe) 사람은 정신 건강 치료법을 찾지 않음

- RHS= "treatment=Yes"로 고정하였을 때,

```
> inspect(asso_us_1_2[1:10])
```

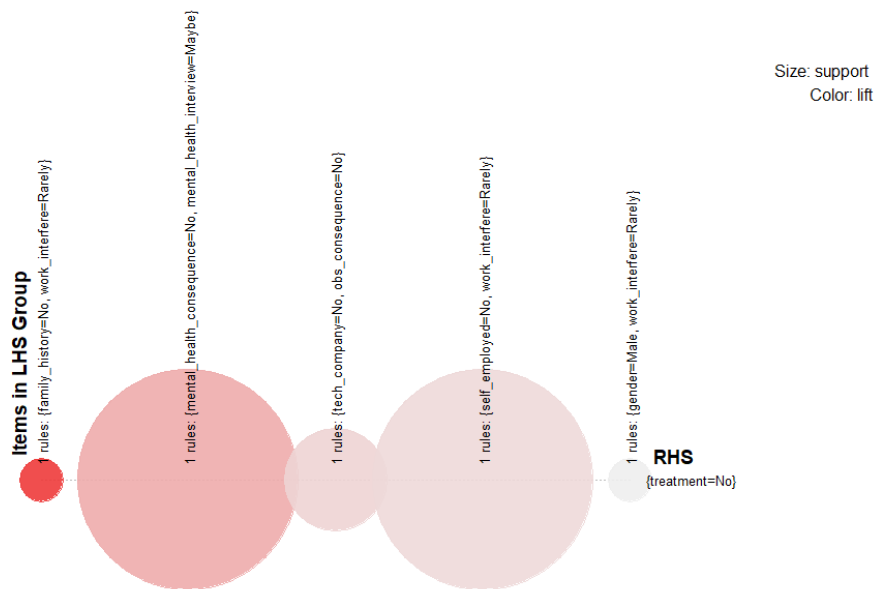
lhs	rhs	support	confidence	lift
[1] {tech_company=yes,vacation=very difficult}	=> {treatment=Yes}	0.05026930	0.7567568	1.124036
[2] {vacation=very difficult,mental_health_interview=No}	=> {treatment=Yes}	0.05745063	0.8000000	1.188267
[3] {state=WA,family_history=yes}	=> {treatment=Yes}	0.05026930	0.8484848	1.260283
[4] {state=WA,benefits=yes}	=> {treatment=Yes}	0.05206463	0.8283714	1.230703
[5] {state=WA,remote_work=No}	=> {treatment=Yes}	0.05206463	0.7073171	1.050602

- ▶ IT 회사에 있고 정신 건강 문제로 병가를 내기 어려운 사람들은 treatment를 찾음
- ▶ 워싱턴 주에 있는 사람들은 가족이 정신 건강 병력이 있으면 정신 건강 치료법을 찾으려는 노력을 보임

country != "United States"이고

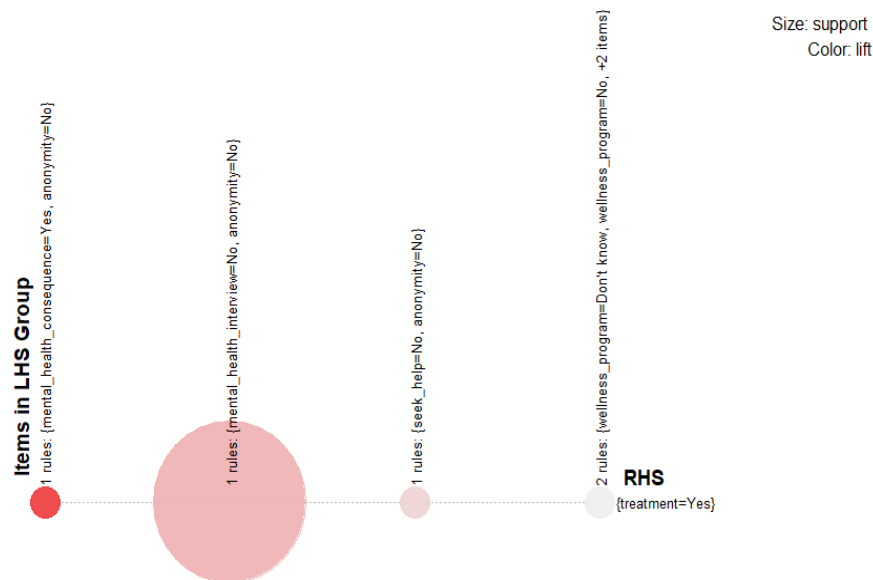
- RHS= "treatment=No"로 고정하였을 때,
- RHS= "treatment=Yes"로 고정하였을 때,

Grouped Matrix for 5 Rules



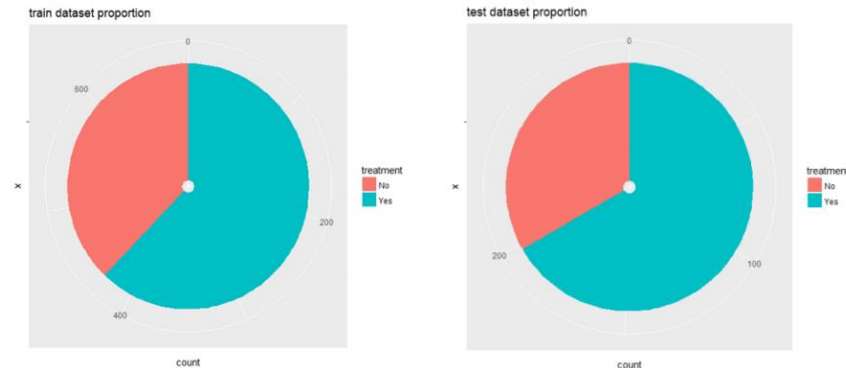
▶ 최대 confidence가 0.5여서 의미 있는 결론이 있다고 보기 힘들

Grouped Matrix for 5 Rules



▶ **익명성이 보호되지 않고**, 잠재적 고용주와 정신건강에 대해 **논의하는 것이 부정적인 결과를 가져오면** 스스로 정신건강에 대한 치료법을 따로 찾는 것으로 나타남

- train data : test data = 7:3 으로 분류



- 로지스틱 회귀분석 & stepwise를 통한 변수 선택 진행

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: treatment
##
## Terms added sequentially (first to last)
##
##
```

	Df	Deviance	Resid. Df	Resid. Dev	Pr(>Chi)
## NULL			390	488.51	
## family_history	1	34.989	389	453.52	3.316e-09 ***
## work interfere	3	120.443	386	333.08	< 2.2e-16 ***
## no_employees	2	1.948	384	331.13	0.377625
## care options	2	20.065	382	311.06	4.394e-05 ***
## wellness_program	2	4.417	380	306.65	0.109878
## anonymity	2	9.253	378	297.39	0.009788 **
## coworkers	1	3.227	377	294.17	0.072429 .
## phys_health_interview	2	4.193	375	289.97	0.122909

```
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- no\_employees, wellness\_program, coworkers, phys\_health\_interview가 유의미하지 않음을 확인
- 해당 변수를 제거하고 로지스틱 회귀와 ANOVA를 진행

- 회귀분석 및 ANOVA 결과

```
## Call:
## glm(formula = treatment ~ family_history + work_interfere + care_options +
##       anonymity, family = binomial, data = data1)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.6218  -0.3536   0.3829   0.6423   2.4666
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)    -0.2479     0.2430  -1.020  0.307589
## family_historyYes  1.0923     0.2363   4.623 3.78e-06 ***
## work_interfere.L  1.8595     0.2411   7.711 1.25e-14 ***
## work_interfere.Q  -1.8637     0.2805  -6.643 3.07e-11 ***
## work_interfere.C  1.4032     0.3194   4.394 1.12e-05 ***
## care_optionsNot sure -0.2522     0.2877  -0.877 0.380731
## care_optionsYes    1.1025     0.2907   3.793 0.000149 ***
## anonymityNo       -0.4515     0.6843  -0.660 0.509356
## anonymityYes       0.5041     0.2800   1.800 0.071791 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 740.63  on 586  degrees of freedom
## Residual deviance: 485.25  on 578  degrees of freedom
## AIC: 503.25
##
## Number of Fisher Scoring iterations: 5
```

```
## Analysis of Deviance Table
##
## Model: binomial, link: logit
##
## Response: treatment
##
## Terms added sequentially (first to last)
##
##              Df Deviance Resid. Df Resid. Dev  Pr(>Chi)
## NULL                                586    740.63
## family_history  1    58.011    585    682.62 2.606e-14 ***
## work_interfere  3   158.681    582    523.94 < 2.2e-16 ***
## care_options    2    34.627    580    489.31 3.026e-08 ***
## anonymity       2     4.060    578    485.25 0.1313
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

▶ 대체적으로 유의미하게 나옴을 확인

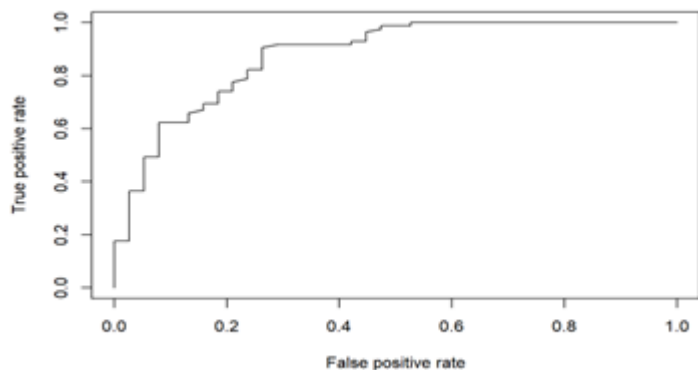
- McFadden R스퀘어로 해당 모델의 Fit 확인

```
pR2(reduced_model2)
```

```
##          llh          llhNull          G2          McFadden          r2ML
## -242.6255660 -370.3153168    255.3795016    0.3448136    0.3527734
##          r2CU
##      0.4921271
```

▶  $R^2 \approx 0.49$ 로 약 49%의 설명력이 있음

- ROC 그래프 및 AUC를 이용한 모델 평가



```
auc <- performance(pr, measure="auc")
auc <- auc@y.values[[1]]
auc
```

```
## [1] 0.8780186
```

해당 모델이 잘  
평가되었음을 확인

- 모델에서 유의한 변수로 선택된 **family\_history, work\_interfere, care\_options, anonymity**가 treatment에 주요한 영향을 끼치는 것으로 파악됨
- care\_options 여부에서 Yes인 경우에도 정신치료를 받기 위한 탐색을 하는 것으로 나타났다. 이를 통해 **표면적으로만 익명성 보장**이 이루어지나 **직원들이 이에 대해 신뢰하고 있지 않을 수도 있다고** 추측할 수 있었음

## 분석 결과

- work\_interfere, family\_history가 treatment 변수에 영향을 많이 줌  
 ▶ 가족 중에 정신질환을 앓은 사람이 있거나 직장에서 정신질환이 업무에 방해를 준다고 느낄 수록 정신 건강 치료법을 찾는 것으로 나타남
- 작업 환경이 부정적인 영향을 끼칠 수록 동료들과 그 경험에 대해 공유를 하여 본인 상태를 점검하는 것으로 나타남  
 ▶ 고용인들이 서로의 정신건강에 대해 말할 수 있는 환경이 마련되면 본인 스스로의 정신건강에 대한 주변 환경을 인식하고 관리할 수 있다고 추측함
- 고용인들의 모국이 미국이 아닐 때, 익명성이 보호되지 않으면 스스로 정신건강에 대한 치료법을 찾아 봄  
 ▶ 고용인이 외국인일 경우 치료를 지원해 준다 하더라도 익명성 보장이 되지 않으면 직원들이 이를 신뢰하지 않고 독립적으로 치료법을 찾는 것으로 나타났다.
- 고용인들의 모국이 미국일 때, IT 회사에 있고 정신 건강 문제로 병가를 내기 어려운 사람들은 스스로가 치료법을 찾는 것으로 나타남  
 ▶ IT회사 특성상 일이 많아 병가를 내는데 회사의 환경이 조성되지 않으면, 고용인이 가지게 되는 부정적 이미지와 신뢰성 하락이 고용주가 제공하는 정신건강 옵션이 있음에도 이용하지 않을 것으로 추측하였다.



## 결론 종합

- ▶ 고용인들 사이의 정신 건강에 대해 말할 수 있는, **집단 상담** 등과 같은 환경을 조성하고
- ▶ 고용인들의 근무 환경을 적절히 개선하여 (고용인들을 대상으로 한) **회사의 이미지 및 신뢰성**을 키우고
- ▶ 정신 건강 복지에 대해 **철저한 익명성**을 보장한다면

고용인들이 보다 편리하게 회사 내에서 본인의 정신 건강 개선에 대한 해결법을 찾을 수 있음





A woman with long brown hair tied in a ponytail, wearing a white cable-knit sweater, is seated at a desk. She is looking down at a laptop, with her right hand resting on her forehead in a gesture of stress or exhaustion. A cup of coffee sits on the desk next to the laptop. The background is a large window with a view of a city skyline. The text "Thank You" is overlaid in the center of the image.

**Thank You**