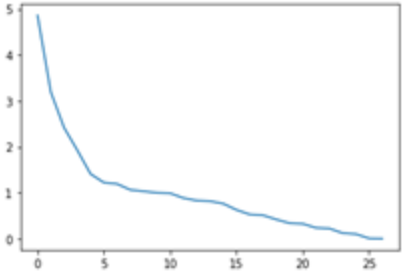
Chevron Project

Ricardo Pérez Castillo A01281092
Angel Treviño Estrada A01411321

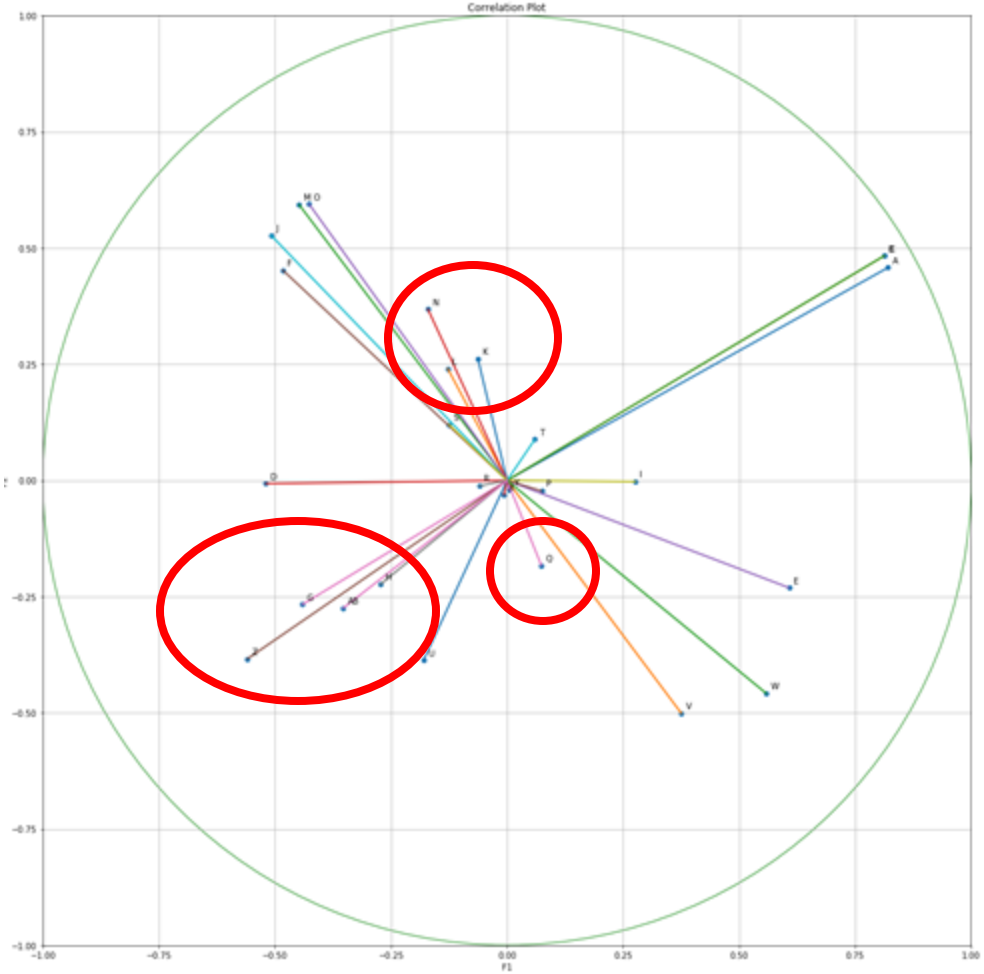
Objective

Given the data, determine under which circumstances (values of the independent variables) the repair/maintenance cost will exceed 60,000 euros.

PCA



4 Principal Components are more that enough



L = time between equipment failures of any kind

Q = Report total repair time

AB = Class (Cost > 60,000)

idx	Var	F1	F2	F3	F4	F5	F6	F7
0	A	0.673835	0.209717	0.014707	0.019271	0.001431	0.007791	0.011289
1	B	0.660307	0.233424	0.018655	0.014124	0.002392	0.010612	0.006866
2	C	0.660167	0.233539	0.018720	0.014084	0.002418	0.010593	0.006890
3	D	0.270083	0.000048	0.016668	0.018392	0.033840	0.121712	0.007394
4	E	0.369236	0.053376	0.066258	0.018516	0.021906	0.010161	0.012885
5	F	0.233503	0.204585	0.179409	0.055943	0.206887	0.001693	0.000317
6	G	0.194450	0.071246	0.035421	0.000159	0.000284	0.133956	0.083853
7	H	0.074353	0.049959	0.111760	0.380902	0.179453	0.039384	0.000026
8	I	0.076096	0.000007	0.173311	0.306405	0.185327	0.067204	0.008661
9	J	0.258727	0.278230	0.073117	0.024867	0.106281	0.047484	0.000275
10	K	0.003939	0.068313	0.045557	0.373383	0.254023	0.000631	0.002068
11	L	0.016319	0.057728	0.482170	0.195429	0.003309	0.015960	0.000815
12	M	0.200570	0.351578	0.241105	0.080002	0.000056	0.001707	0.001189
13	N	0.029079	0.135656	0.244057	0.015693	0.242652	0.013825	0.003312

idx	Var	F1	F2	F3	F4	F5	F6	F7
14	O	0.181863	0.353273	0.197621	0.092440	0.002013	0.001808	0.003060
15	P	0.005559	0.000517	0.001363	0.000084	0.003423	0.010115	0.000978
16	Q	0.005456	0.033459	0.006586	0.024752	0.049028	0.114745	0.033874
17	R	0.003598	0.000139	0.073727	0.105103	0.028114	0.306376	0.006028
18	S	0.015715	0.014488	0.028994	0.014920	0.024648	0.140497	0.126563
19	T	0.003575	0.007773	0.075398	0.000851	0.008868	0.000821	0.361333
20	U	0.032477	0.150008	0.055276	0.005478	0.001481	0.029831	0.146425
21	V	0.140792	0.250815	0.124704	0.105132	0.018663	0.001124	0.002373
22	W	0.311846	0.210003	0.105250	0.038083	0.022821	0.010110	0.001404
23	X	0.000043	0.000940	0.000026	0.000432	0.002841	0.000180	0.002448
24	Y	0.000013	0.000429	0.000014	0.005526	0.001957	0.060013	0.016222
25	Z	0.312994	0.147365	0.025907	0.013405	0.000115	0.057359	0.012363
26	AB	0.124597	0.075443	0.000084	0.000492	0.002687	0.004517	0.331894

PCA

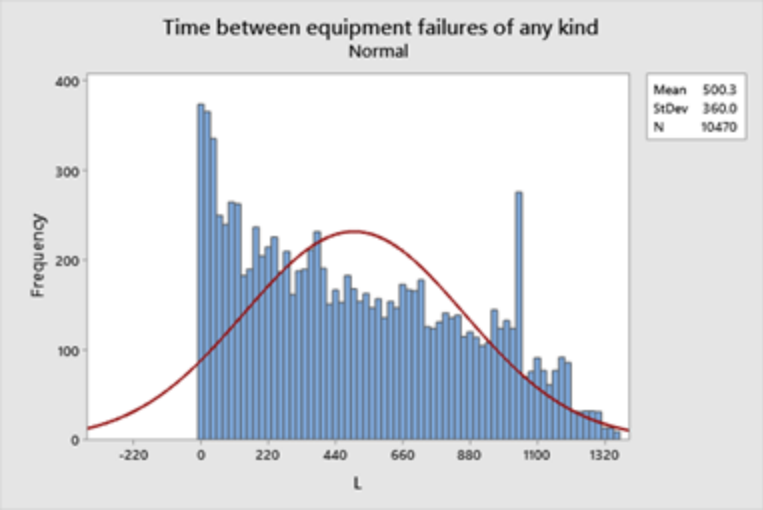
According to the PCA,

Time between equipment failure of any kind (L) belongs to the 3th componente (F3) and seems to be related to tiempo entre equipo fallo (N), tiempo entre equipos assist type (M), tipo de asistencia (F), tipo fracción (I), and tipo equipo (H) in that order

Total report time (Q) belongs to the 6th component (F6) and seems to be related to inmersion (R) , acompañante (S), hay materiales (Y), cantidad materiales(Z) and red vertido (U) in that order

Cost (AB) belongs to the 7th component (F7) and seems to be related to trabajo finalizado (T), red vertido (U), acompañante (S), tiempo trabajado (Q) and hay material (Y) in that order

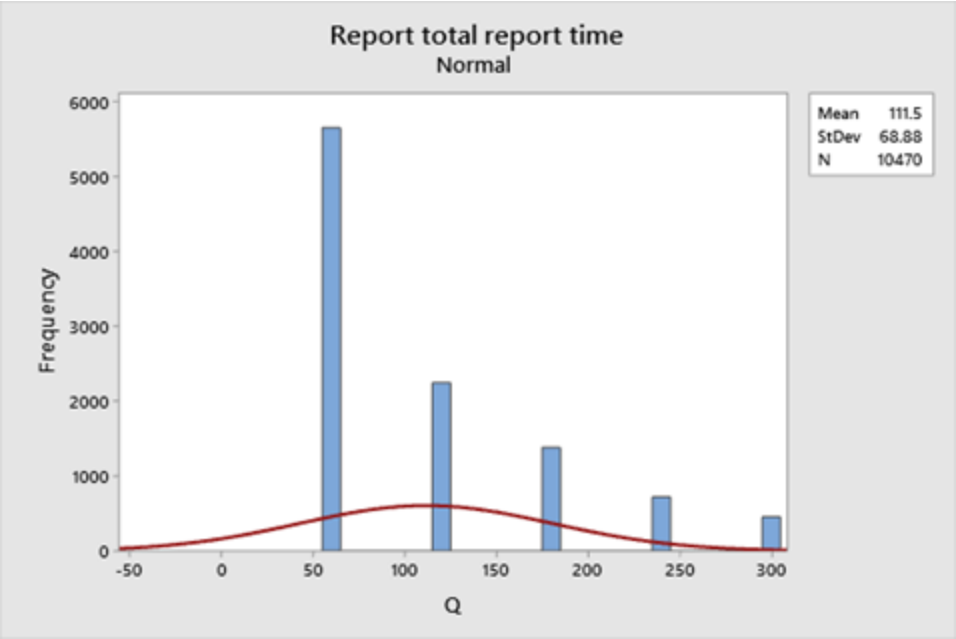
Statistics about time between equipment failures of any kind (L)



Statistics

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
L	10470	0	500.34	3.52	359.98	1.00	183.00	449.00	786.00	1365.00

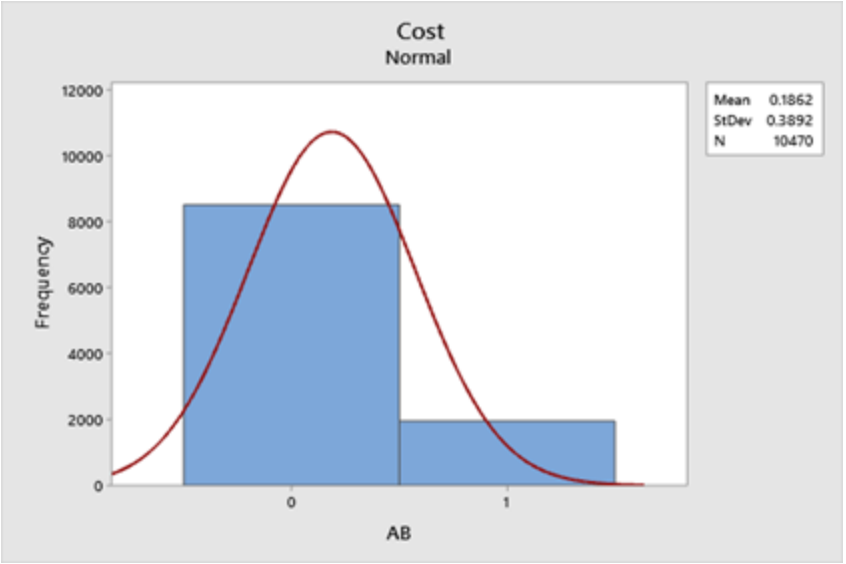
Statistics total report time (Q)



Statistics

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
Q	10470	0	111.54	0.673	68.88	60.00	60.00	60.00	120.00	300.00

Statistics Cost (AB)



Statistics

Variable	N	N*	Mean	SE Mean	StDev	Minimum	Q1	Median	Q3	Maximum
AB	10470	0	0.18615	0.00380	0.38925	0.00000	0.00000	0.00000	0.00000	1.00000

Machine Learning (Deep Learning) to determine Costo (AB) > 60,000 euros

Test Size	Nodes	Max Iter	Likelihood Ratio	Accuracy	F1 measure
0.20	10,10,10,10	1500	135.3333333333337	0.9985673352435	0.996319018404908

Variables used according to PCA

```
In [12]: target_names = ['class 0', 'class 1']
         print(classification_report(y_test, y_pred, labels=[0,1]))
```

	precision	recall	f1-score	support
0	1.00	1.00	1.00	1685
1	1.00	0.99	1.00	409
accuracy			1.00	2094
macro avg	1.00	1.00	1.00	2094
weighted avg	1.00	1.00	1.00	2094