



# **ESCUELA POLITECNICA NACIONAL**



## **FACULTAD DE INGENIERÍA DE SISTEMAS**

### **ISWD433-FUNDAMENTOS DE SISTEMAS DE LA INFORMACIÓN**

**TÍTULO: Limpieza de datos**

**Alumnos: Fernando Huilca Villagómez**

**PROFESOR: Andrés Larco**

**FECHA DE ENTREGA: 27-10-2024**

A continuación, se presenta el informe de los errores que se encontraron en la base de datos proporcionados, para ello se usó un pequeño software computacional en el lenguaje de Java.

### Errores de Cédula:

- Posiciones en las que se presenta errores de campos nulos:  
1,9,17,20,31,41,43,45,58,59,60,61,62,63,70,75,80,94,98,107,110,111,112,113,114,115,116,124,135,138,150,151,153,157,158,159,160,161,162,163,164,165,166,172,173,209,213,217,224,225,226,227,229,230,236,237,246,254,256,261,264,265,266,267,268,269,270,271,272,273,278,279,280,281,282,283,284,285,286,287,291,302,304,308,320,329,332,334,335,345,346,347,348,349,360,369,382,389,404,405,406,407,408,409,410,411,412,413,414,415,416,421,422,428,432,434,440,446,454,455,458,464,467,469,470,471,472,473,474,475,476,477,478,479,480,482,483,485,490,513,516,523,538,551,552,553,554,555,556,564,565,570,577,581,590,597,598,599,600,601,618,636,646,647,648,649,650,651,652,653,663,667
- Posiciones en las que se presenta campos duplicados:  
380,574,661,662,664,665,666,668,669,670,671
- Posiciones en las que se presenta nueve dígitos (posible error):  
2,4,5,6,7,8,11,12,13,14,15,18,19,21,23,24,25,26,27,28,29,30,32,33,34,35,36,37,39,40,42,44,46,47,48,49,50,51,53,54,55,56,57,64,65,66,67,68,71,72,73,74,76,77,78,79,81,82,83,84,85,86,87,88,89,90,91,92,93,95,96,97,99,100,101,103,104,105,106,108,109,117,118,121,122,123,125,126,127,128,129,130,132,133,134,136,137,139,140,142,143,144,145,147,149,152,154,155,156,167,168,169,170,171,174,175,176,177,178,180,181,183,185,186,187,188,189,190,191,192,193,194,195,196,197,198,199,200,201,202,203,204,205,206,207,208,210,211,212,214,215,216,218,220,221,222,223,228,231,232,233,235,238,239,240,241,243,244,245,247,248,249,250,251,252,253,255,258,260,263,275,276,289,292,293,294,295,298,299,300,301,305,306,310,311,312,313,314,315,316,317,318,321,322,323,324,325,326,327,330,331,333,336,337,338,339,342,343,344,350,351,352,353,354,355,356,358,359,362,363,364,365,366,367,368,370,371,373,374,375,376,377,378,381,383,384,386,390,391,392,394,396,397,398,399,400,401,402,403,418,419,420,423,424,425,426,427,429,430,431,433,435,436,437,438,439,441,442,444,445,447,448,449,451,452,453,456,457,459,460,461,462,465,466,468,481,484,486,488,489,491,492,493,494,495,496,497,498,499,501,502,503,504,505,506,507,509,510,511,512,515,517,518,519,520,521,522,524,525,526,528,529,530,531,532,533,535,536,537,539,540,541,542,543,545,546,547,549,550,558,559,560,561,562,563,566,567,568,571,575,576,578,579,580,582,583,584,585,586,587,588,589,591,592,593,594,595,602,603,604,605,606,607,608,609,611,612,613,614,615,616,619,620,621,622,623,624,625,626,627,628,629,630,632,634,635,638,639,640,641,642,643,644,645,655,656,657,658,659
- Posiciones en las que se presenta menos de nueve dígitos:  
No hay
- Posiciones en las que se presenta error de provincia en la cédula:  
234,393,617
- Posiciones en las que se presenta errores de validación del tercer dígito  
No hay
- Posiciones en las que se presenta errores en la validación de los dígitos consecutivos:  
no hay

- Posiciones en las que se presenta errores en la validación del dígito validador:  
16,22,69,102,119,120,141,146,182,184,219,234,242,257,259,274,277,288,297,303,30  
9,319,340,372,379,387,388,393,395,450,463,500,508,514,548,557,569,572,596,610,6  
17,631,637,654,660

**RESULTADO FINAL:** Cantidad de errores en cédulas: 650

#### **Errores de Teléfono:**

- Posiciones en las que se presenta un número diferente a 7 dígitos:  
9,11,13,25,46,151,159,168,263,273,370,446,535,591,662,667,668,669
- Posiciones con campos vacíos o nulos:  
9,168,446,535,668

**RESULTADO FINAL:** Cantidad de errores en teléfonos: 18

#### **Errores en fechas y edades:**

- Posiciones donde la fecha y edad no coinciden:  
205,208,296

**RESULTADO FINAL:** Cantidad de errores en fechas y edades: 3

- Posiciones de campos nulos:  
No hay nulos en ningún campo de edad ni fecha

#### **Errores en calificaciones:**

- Posiciones con campos nulos:  
23,73,122,167,245,297,396,474,520,613,627
- Posiciones fuera de rango:  
7,65,181,355,361,464,507,560,587

**RESULTADO FINAL:** Cantidad de errores en calificaciones: 20

#### **Errores en nombres:**

- Posiciones con campos vacíos:  
17,134,298,450

**RESULTADO FINAL:** Cantidad de errores en nombres vacíos: 4

#### **Errores en apellidos:**

- Posiciones con campos vacíos:  
17,134,298,450

**RESULTADO FINAL:** Cantidad de errores en apellidos vacíos: 4

#### **Errores en correos electrónicos:**

- Posiciones con formato incorrecto ([nombre.apellido@universidad.edu.ec](mailto:nombre.apellido@universidad.edu.ec))  
2,6,7,10,11,13,17,20,24,26,31,32,37,41,44,51,59,61,72,73,77,80,87,95,98,102,110,124,125,12  
8,134,139,141,150,151,152,153,157,158,159,161,164,165,166,177,178,186,209,213,220,223,2  
24,230,242,245,246,261,270,275,278,279,280,281,286,287,298,302,304,306,326,330,332,335,

**RESULTADO FINAL:** Cantidad de errores en correos: 127

Cedula:			
Campos Duplicados color:			
Nueve dígitos posible error:			
Error de provincia color:			
Error dígito validador color:			
Campos nulos color para identificarlos:			
Cantidad de errores en cédulas: 650			
Telefono:			
Distinto a 7 dígitos color:			
Campos nulos color para identificarlos:			
Cantidad de errores en teléfonos: 18			
Edad y fecha:			
No coinciden color:			
Cantidad de errores en fechas y edades: 3			

Nombre y apellidos:			
Campos nulos color para identificarlos:			
Cantidad de errores en nombres vacíos: 4			
Calificaciones:			
Fuera de rango sobre 20:			
Campos nulos color para identificarlos:			
Cantidad de errores en calificaciones: 20			
Correos:			
Formato incorrecto			
Cantidad de errores en correos: 127			