VIET NAM NATIONAL UNIVERSITY HO CHI MINH CITY
HO CHI MINH CITY UNIVERSITY OF TECHNOLOGY
FACULTY OF COMPUTER SCIENCE AND ENGINEERING

Course name h

Report type h

Report title h

Advisor: Advisor h

HO CHI MINH CITY, APRIL 2023

Contents

1 Abstract

2 Background

3 Data description

3.1 Importing data

For the first part of our project, we need to select a suitable dataset for us to analyze, as we are computer science students, we have decided to select a CPU data set, the data of which can be found here:
The data give us information about 2283 CPU and 45 of their feature which include:

- Product_Collection: tell us which type of series the core belongs to.
- Vertical_Segment: show what kind of system the CPU was designed for (embedded, mobile, desktop, or sever).
- Processor_Number : process ID.
- Status: show the status of the CPU (announce, launched, end of life, end of support).
- Launch_Date: The date the product was first introduced.
- Lithography: refers to the semiconductor technology used to manufacture an integrated circuit, and is reported in nanometers (nm), indicative of the size of features built on the semiconductor.
- Recommended_Customer_Price: recommended customer price.
- nb_of_Cores: total number of cores in a proccessor.
- nb_of_Threads: total number of thread in a processor.
- Processor_Base_Frequency: Describes the rate at which the processor's transistors open and close.
- Max_Turbo_Frequency: The maximum single core frequency at which the processor is capable of operating using Intel® Turbo Boost Technology.
- Cache: CPU Cache is an area of fast memory located on the processor.
- Bus_Speed: refers to how much data can move across the bus simultaneously.
- TDP(thermal design power): Represents the average power, in watts, the processor dissipates when operating at Base Frequency with all cores.
- Embedded_Options_Available: is it allow to be embedded system
- Conflict_Free: Defined by the U.S. Securities and Exchange Commission rules to mean products that do not contain conflict minerals (tin, tantalum, tungsten).
- Max_Memory_Size: The maximum memory capacity supported by the processor.

- Memory_Types: Single Channel, Dual Channel, Triple Channel, and Flex Mode.The maximum memory capacity supported by the processor.
- Max_nb_of_Memory_Channels: The number of memory channels refers to the bandwidth operation for real world application.
- Max_Memory_Bandwidth: The maximum rate at which data can be read from or stored into a semiconductor memory by the processor (in GB/s).
- ECC_Memory_Supported: ECC memory is a type of system memory that can detect and correct common kinds of internal data corruption.
- Processor_Graphics: integrated graphics processing unit (GPU) that is built into some of Intel's processors.
- Graphics_Base_Frequency: The rated/guaranteed graphics render clock frequency in MHz.
- Graphics_Max_Dynamic_Frequency: The maximum opportunistic graphics render clock frequency (in MHz) that can be supported using Intel HD Graphics with Frequency feature.
- Graphics_Video_Max_Memory: The maximum amount of memory accessible to processor graphics. Processor graphics operates on the same physical memory as the CPU (subject to OS, driver, and other system limitations).
- Graphics_Output: Graphics Output defines the interfaces available to communicate with display devices.
- Support_4k: indicates the product's support of 4K
- Max_Resolution_HDMI: the maximum resolution supported by the processor via the HDMI interface (24bits per pixel & 60Hz). System or device display resolution is dependent on multiple system design factors; actual resolution may be lower on your system.
- Max_Resolution_DP: The maximum resolution supported by the processor via the DP interface (24bits per pixel & 60Hz). System or device display resolution is dependent on multiple system design factors.
- Max_Resolution_eDP_Integrated_Flat_Panel
- DirectX_Support: Indicates support for a specific version of DirectX, a Microsoft collection of APIs for handling multimedia compute tasks.
- OpenGL_Support: Indicates support for OpenGL, a cross-language, multi-platform API for rendering 2D and 3D vector graphics.
- PCI_Express_Revision: The PCIe version supported by the processor.
- PCI_Express_Configurations_: The available PCIe lane configurations that can be used to link the PCH PCIe lanes to PCIe devices.

- T : The maximum temperature allowed on the chip.
- Max_nb_of_PCI_Express_Lanes: maximum number of PCI Express Lanes that are supported.
- Intel_Hyper_Threading_Technology_: Delivers two processing threads per physical core. Highly threaded applications can get more work done in parallel, completing tasks sooner.
- Intel_Virtualization_Technology_VTx_: Allows one hardware platform to function as multiple "virtual" platforms. It offers improved manageability by limiting downtime and maintaining productivity by isolating computing activities into separate partitions.
- Intel_64_: Delivers 64-bit computing on server, workstation, desktop and mobile platforms when combined with supporting software. Intel 64 architecture improves performance by allowing systems to address more than 4 GB of both virtual and physical memory.
- Instruction_Set: Which instrution set the CPU use.
- Instruction_Set_Extensions : Instruction set extension
- Idle_States: Used to save power when the processor is idle.
- Thermal_Monitoring_Technologies: Protects the processor package and the system from thermal failure through several thermal management features.
- Secure_Key: The CPU is supported with secure key or not.
- Execute_Disable_Bit: Hardware-based security feature that can reduce exposure to viruses and malicious code attacks.

For coding the data, our team use a wide range of package which include

- rio: for basic import of data.
- ggplot2: for dealing with plot formats.
- zoo: for dealing with the year quarter format
- ggpubr: to further customize the plot of ggplot

```
# pacman::p_load(rio,     # for dealing with basic import export
               ggplot2,  # for dealing with plot formats
               zoo,      # for dealing with year quarter formats
               ggpubr)   # customize ggplot
```

3.2 Data preprocessing

3.3 Data cleaning

4 Data clarification

5 Data analysis

5.1 Analysis of Variance (ANOVA)

5.2 Regression

6 Conclusion