

வணிகம்

Published: August 28, 2014 10:00 IST Updated: August 28, 2014 12:15 IST

பிக் டேட்டாவின் குணாதிசயங்கள் என்ன?

ரவி சின்னதம்பி



பிக் டேட்டா என்றால் என்ன? டேட்டாவிலேயே பெரிய அப்பாடக்கரா? வண்டிவண்டியாய் டேட்டா இருந்தால் மட்டுமே அதனை பிக் டேட்டா என்று சொல்லிவிட முடியாது. பிக் டேட்டாவிற்கு நான்கு குணாதிசயங்கள் உண்டு. எளிதில் புரிவதற்கான உதாரணங்கள் சிலவற்றை தந்துள்ளேன். டேட்டா கோட்பாடுகள் இடிக்கின்றது என யாரும் வரிந்துகட்டிக்கொண்டு சண்டைக்கு வரக்கூடாது.

எது பிக் டேட்டா ?

சென்னையில் வாழும் மனிதர்களின் பெயர்களை மட்டும் டேட்டா பேஸில் ஏற்றிவிட்டால் அது பிக்-டேட்டாவாகிவிடாது. பரபரப்பான சென்ட்ரல் ரயில் நிலையத்தின் வாசலில் ஒரு ஸ்கேனர் வைத்து எத்தனை பேர் ரயில் நிலையத்தின் உள்ளேயும் வெளியேயும் சென்றார்கள் என்று கவுண்டிங் செய்து பதிந்தால் அது பிக்-டேட்டாவாகாது.

அதிகாலையில் புறப்படத்தயாராக இருக்கும் ஒரு ரயிலில் எத்தனை பேர் செய்தித்தாள் விற்கும் பையனிடம் நாளிதழ்களை வாங்கினார்கள் என்று பதிந்து வைத்தால் அதுவும் பிக் டேட்டாவாகாது. அப்படி நாளிதழ்களை வாங்கினவர்கள் அனைவருமே அந்த நாளிதழ்களின் சந்தாதாரர்களாக இருப்பார்கள் என்று நாமாகவே எழுதிவைத்துக்கொண்டால் அதுவும் பிக் டேட்டாவாகாது.

பிக் டேட்டா என்ற தகுதியைப் பெற முக்கியமான நான்கு குணாதிசயங்கள் வேண்டும். அதிகமான பதிவுகள், அதிவேகமான பதிவுகள், பலவிதமான பதிவுகள், உண்மையான உபயோகத்தன்மையுள்ள பதிவுகள் என்ற நான்கும் பிக் டேட்டாவை விவரிக்கும் குணாதிசயங்கள். [ஆங்கிலத்தில் 4 V's என்று சொல்லும் - Volume, Velocity, Variety and Veracity]. இந்த நான்கு குணங்களைக் கொண்ட பிக் டேட்டாவை பிசினஸ் உபயோகத்திற்கு கொண்டுவருவது மற்றுமொரு 'V' என்ற Value எனும் ஐந்தாவது V.

மதிப்புதான் உயிர்

எல்லா குணங்களும் இருந்தும் அந்த டேட்டாவில் மதிப்பு இல்லாவிட்டால் என்ன உபயோகம் இருக்கப்போகின்றது? பிசினஸிற்கு வேல்யுதானே உயிர். இந்த நான்கு குணங்களைக்கொண்ட பிக் டேட்டாவை வைத்துக்கொண்டு எந்தவகையான பிசினஸ் மதிப்புகளை கூட்ட முடியும் என்பதுதான் அது.

இந்தக் குணாதிசயங்களை வைத்துக்கொண்டு மேலே சொன்ன உதாரணங்களை அலசுவோம். சென்னையில் வாழும் நபர்களின் பெயரை தொடர்ந்து முகவரி, கைபேசி எண், வீட்டிலிருக்கும் எரிவாயு கணக்ஷன் கம்பெனியில் ஆரம்பித்து டிஷ் டிவி, போன், கம்ப்யூட்டர், சோபா செட் வாங்கிய கடை, ரெகுலராய் நகை வாங்கும் கடை, பிரியப்பட்டு செல்லும் சினிமா தியேட்டர்,

ரெஸ்ட்டாரெண்ட், துணிக்கடை, வாங்கும் பணியனின் பிராண்ட், வைத்திருக்கும் டீ விலர், வாரா வாரம் போகும் கோவில், மருந்துவாங்கும் கடை என ஜாதகமே இருந்தால் அது பிக் டேட்டாவிற்கு முதல் தகுதியை பெற்றுவிட்ட ஒன்று.

சென்ட்ரல் ரயில் நிலையத்திற்கு உள்ளேயும் வெளியேயும் சென்றவர்களின் பயணத்திட்டம் எத்தனை நாட்கள், கையில் இருக்கும் லக்கேஜின் (பெட்டி/கைப்பை) பிராண்ட், எங்கே தங்கினார் / தங்கப்போகின்றார், டாக்ஸி பிடிப்பாரா? ஏற்கனவே கார் காத்திருக்கின்றதா? ரிட்டர்ன் ஜர்னி டிக்கெட் இருக்கின்றதா? அடிக்கடி சென்னை வருபவரா? மாமாங்கத்துக்கும் ஒரு முறையா? என பயணத்தை தாண்டிய பல டேட்டாக்கள் அந்த நபர் சென்ட்ரல் ரயில் நிலைய வாயிலை கடக்கும் நொடிப்பொழுதில் பதிவு செய்ய முடிந்தால் அதுவும் பிக்-டேட்டாவுக்கு தகுதிபெறும் மற்றொரு விஷயமாகும்.

புறப்படத்தயாராக இருக்கும் ரயில் வண்டியில் பேப்பர் வாங்கினவர்கள் எத்தனை பேர் என்ற விவரத்தைத் தாண்டி என்ன கோச், நாளிதழ், வாராந்திர இதழ்கள், புத்தகம், வாட்டர் பாட்டில், டிபன், கடலை மிட்டாய், வறுகடலை, டூத் பேஸ்ட், வாழைப்பழம் போன்றவை, தனிநபர் பயணமா, குடும்பத்துடன், அலுவலக பணியாளர்களுடன் பயணமா?, ரயிலில் வாங்கிய நாளிதழ்/வாராந்திர இதழ்களை அவர் தொடர்ந்து வாங்குகின்றாரா? அல்லது வேறு நாளிதழ்/புத்தகங்களை வாங்குகின்றாரா? போன்ற இன்னபிற டேட்டாக்களும் பதிவு செய்யப்பட்டால் அது பிக்-டேட்டாவின் குணாதிசயத்தை சற்றே உள்ளடக்கியது எனலாம்.

இந்த உதாரண டேட்டா கலெக்ஷனில் வரும் டேட்டாவினால் ஏதாவது ஒரு முக்கிய உபயோகமும் இருக்கவேண்டும். பிறந்த ஊர், முதலில் படம் பார்த்த டெண்டு கொட்டகையின் பெயர் எனத் தேவையில்லாத டேட்டாக்கள் பெருமளவில் இருந்தால் அது பிக் டேட்டாவாகாது.

வால்ட்ரம் ரகசியம்

அட! இதெல்லாம்தான் டேட்டாவா? இதையெல்லாம் வைத்துக்கொண்டு என்ன செய்ய முடியும் என்ற எண்ணம் உங்கள் மனதில் தோன்றலாம். இவை டேட்டாவின் உதாரணங்கள் மட்டுமே. டேட்டாக்கள் தொடர்ச்சியாகவும் வேகமாகவும் அலசப்படும் போதுதான் பெரியபெரிய உண்மைகள் புலப்படுகின்றன. உதாரணத்திற்கு, நீங்கள் கிரெடிட் கார்ட் வைத்திருக்கின்றீர்கள்.

அடிக்கடி வெளியூர் செல்லும் வேலை. உங்கள் கம்பெனி தங்கும் செலவு நாளுக்கு 1,200 ரூபாய் வரை தரும். எப்போதும் அதற்கு உட்பட்டே உங்கள் செலவு இருக்கும். கிரெடிட் கார்டிலேயே அதை எப்போதும் கொடுக்கும் பழக்கம் கொண்டவர் நீங்கள். ஒரு நாள் உங்கள் கிரெடிட் கார்டு தொலைந்துபோகின்றது. நீங்கள் அதை உணருமுன் ஒரு பைவ் ஸ்டார் ஓட்டல் ரெஸ்ட்டா

ரென்டில் அந்த கார்டு உபயோகிக்கப்படுகின்றது. கார்டு ஸ்வைப்பாகி உங்களுக்கு தகவல் வருவதற்கு முன்னாலேயே கிரெடிட் கார்டு கம்பெனியின் சர்வரில் ஓடும் அனலிடிக்ஸ் சாப்ட்வேர் அட! நம்ம ஆளு இதுவரை இந்த மாதிரி காஸ்ட்லி ஓட்டலுக்கு போனதேயில்லையே என்று ஸ்வைப் செய்த ஹோட்டலுக்கு "கார்ட் கொண்டு வந்த ஆளை எதற்கும் சரிபார்க்கவும்" என்று ஒரு அலர்ட் கொடுத்தால் திருடன் பிடிபடுவான் இல்லையா?!

வெறுமனே செலவுகளை கணக்கு வைத்து வசூலைச் செய்யாமல் பல்வேறு அனலிடிக்ஸ்களை கிரெடிட் கார்டு நிறுவனங்கள் உபயோகிக்கின்றன. எப்போதும் இல்லாத அளவில் அடிக்கடி பாரில் உங்கள் கிரெடிட் கார்ட் உபயோகிக்கப்படுகின்றதா? ஏதோ சோகத்திலோ அல்லது ஊதாரித்தனத்திலோ திரியும் நம்ம ஆளு பணம் கட்டாமல் போவதற்கு வாய்ப்பு அதிகம் உள்ளது என்று கிரெடிட் கார்ட் நிறுவனம் அனலிடிக்ஸ் மூலம் புரிந்துகொள்கின்றது.

அடிக்கடி என் அக்கவுண்டில் ஏன் பில் அதிகமாக வருகின்றது, குறையர மாதிரியே தெரியலயே, உங்க கம்பெனியில்தாங்க இவ்வளவு காசாகுது என்று கஸ்டமர் சர்வீஸிற்கு போன் செய்து பேசும் போஸ்டு பெய்டு கஸ்டமர்களில், இத்தனை பேர் நம்பர் போர்ட்டபிலிட்டியில் நம்மை விட்டு வெளியே போய்விடும் வாய்ப்பு இருக்கின்றது என்று கண்டறிந்துகொண்டுவிடுகின்றன மொபைல் டெலிகாம் நிறுவனங்கள். ஒத்தையாள்தானே! போனால் போகட்டும் போடா என்று

இருந்துவிட முடியாது என்று சொல்கின்றது அனலிடிக்ஸ்.

ஒருவர் செல்போன் கனெக்ஷனை மற்றோரு கம்பெனிக்கு மாற்றினால் அவருடைய போனில் கான்டாக்ட்ஸில் இருப்பவர்களில் அதிக பட்சம் ஏழுபேர் அவர் போன புதிய போன் கம்பெனிக்குப் போய்விடுகின்றார்கள் என்று எச்சரிக்கை மணியடிப்பது அனலிடிக்ஸ். ஆயிரத்தில் ஒருத்தர் போனால் போகட்டும் என்று இருந்தால் லட்சத்திற்கு எழுநூறு பேர் போய்விடுவார்கள் இல்லையா? வால்ட்யூம்தான் டேட்டாவின் ரகசியம் என்பது புரிகின்றதா?

cravi@seyyone.com

Keywords: [வால்ட்யூம்](#), [மதிப்பு](#), [பிக் டேட்டா](#), [டேட்டா கோட்பாடுகள்](#), [டேட்டா பேஸ்](#), [தகவல் களஞ்சியம்](#), [தகவல் சேகரிப்பு](#)

Topics:

[வணிகம்](#)